じっくり勉強すれば身につく統計入門

今回の「じっくり勉強すれば身につく統計解析」は，サイエンティスト社から刊行されている「医薬品開発のための統計解析」の範囲から発展して，打ち切りデータがある場合の回帰分析を取り上げる．

正規分布を仮定した打ち切りデータの回帰分析入門

高橋　行雄　（Biostat 研究所）

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正確な平均値及び標準偏差を推定することができる．あるデータのその正規分布で占める位置は，対応する確率密度とみなすことができ，全データの確率密度の積が最大になる平均値及び標準偏差が最も適合性が高い正規分布と推定できる．また，各用量群の標準偏差が同じであると仮定できれば，正確な回帰分析を行うこともできる．一方，右側打ち切りのある場合は打ち切り点以上の確率密度の面積に当たる上側確率とみなすことで，確率密度と同じに扱うことができる．Excel の各関数及びソルバーを用いて，打ち切りがある場合の平均値と標準偏差及び回帰分析のパラメータを推定する手順を示し，考え方について丁寧に解説する．

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

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. 正規分布確の率密度の活用 24
6. 正規分布の上側確率の活用 34
7. 打ち切りを含む回帰分析 45
8. JMPの出力を Excel で活用 57
まとめ 64

1．打ち切りデータの実例

発毛までの日数

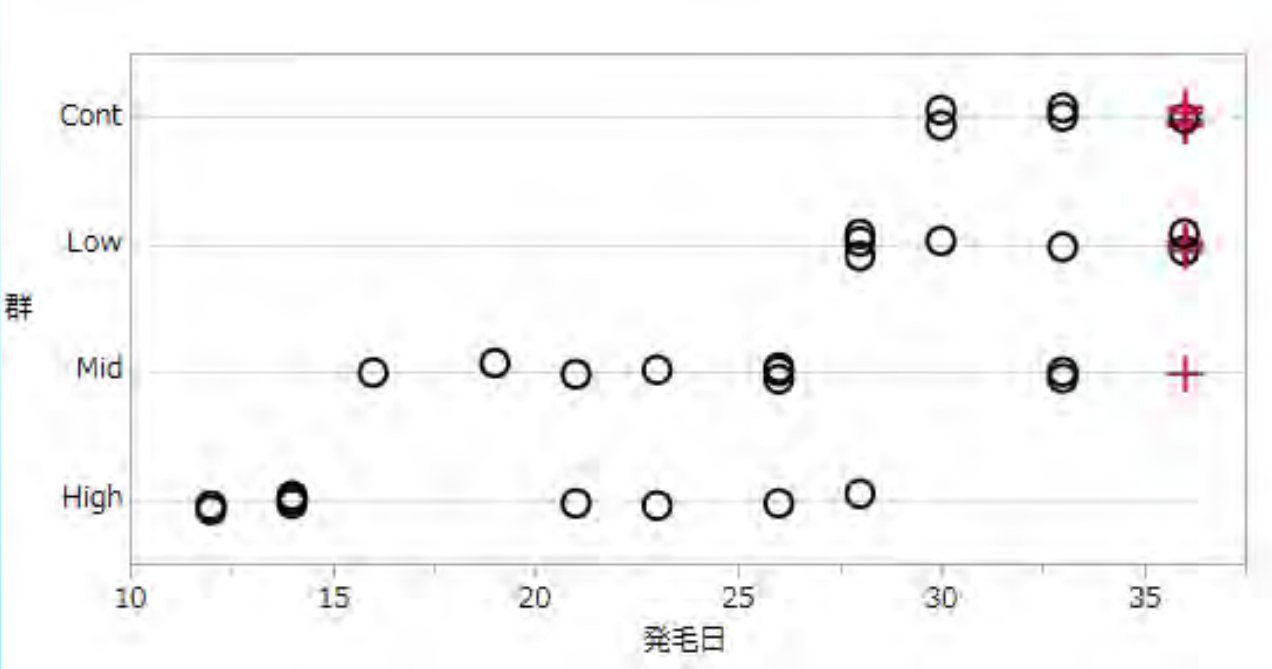
- 実験の概要
  - 発毛までの日数を測定
  - 試験期間は，36日間
  - 対照群と3用量の群間比較試験
  - 1群当たり10匹
- 反応変数
  - 36日目までに発毛した場合の日数
  - 36日に発毛していなければ，36日（打ち切りあり）

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

発毛までの日数の散布図



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.

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.

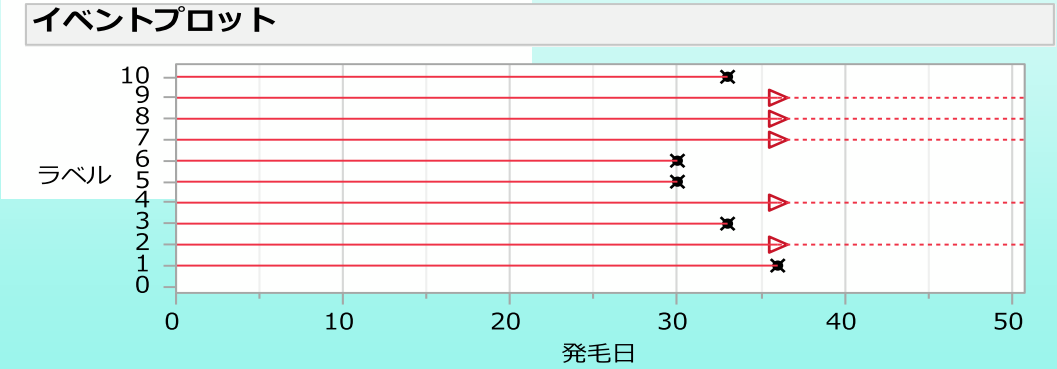
群ごとの平均値と標準偏差

- 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し知れない．

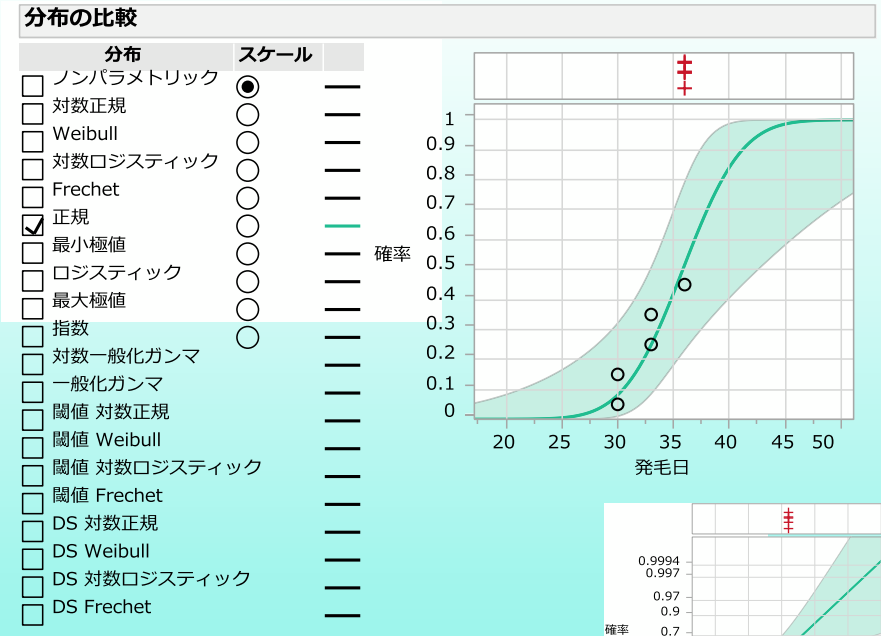
2．打ち切りを考慮した統計量

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



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.

対照群の平均値と標準偏差

パラメトリック推定 – 正規

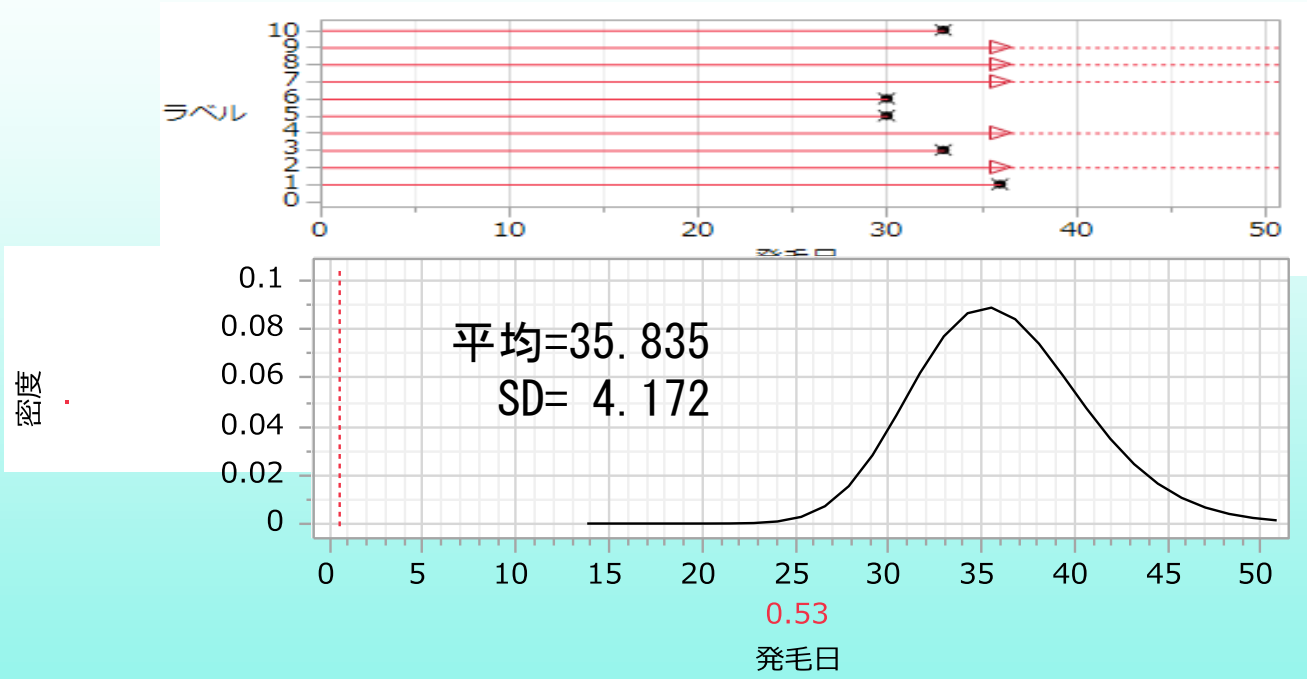
パラメータ	推定値	標準誤差	下側95%	上側95%
位置	35.835	1.623	32.654	39.015
尺度	4.172	1.484	1.264	7.080
平均	35.835	1.623	32.654	39.015

規準	
(-2)*対数尤度	35.562677
AICc	41.276963
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+

正規分布・確率密度



３．段階的な学習

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．

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）では，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.

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.

4. 最小2乗法の復習

2013/10/26　第13回定例会 午前セミナー

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

回帰分析の一歩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，回帰直線が求まれば 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，使う機会は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．

http://www.scientist-press.com/edit_html/green/seminar7.pdf

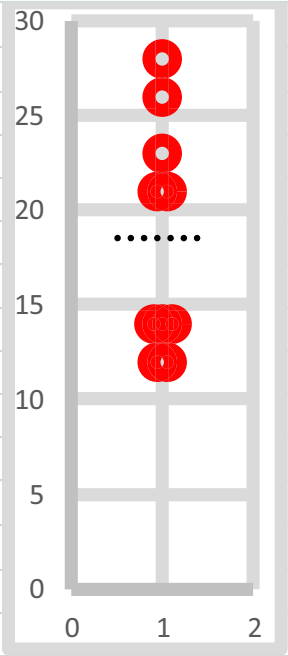
偏差平方和

- 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^ としたときに，μ^ 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^ を推定する．
- 最小 2 乗推定量 μ^ 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^ は，ミュー・ハットと読む．

高用量群：最小2 乗平均

		推定値	偏差	偏差平方
拡散x	y	y^	e	e^2
0.95	12	18.5	−6.5	42.250
1.05	12	18.5	−6.5	42.250
0.9	14	18.5	−4.5	20.250
1	14	18.5	−4.5	20.250
1.1	14	18.5	−4.5	20.250
0.95	21	18.5	2.5	6.250
1.05	21	18.5	2.5	6.250
1	23	18.5	4.5	20.250
1	26	18.5	7.5	56.250
1	28	18.5	9.5	90.250
	偏差平方和（最小化）			324.500
			不偏分散	36.056
			標準偏差	6.005



μ^
18.5

動かすセル

平均値の線	
18.5	0.5
18.5	1.5
y	x

平均を変化させた結果

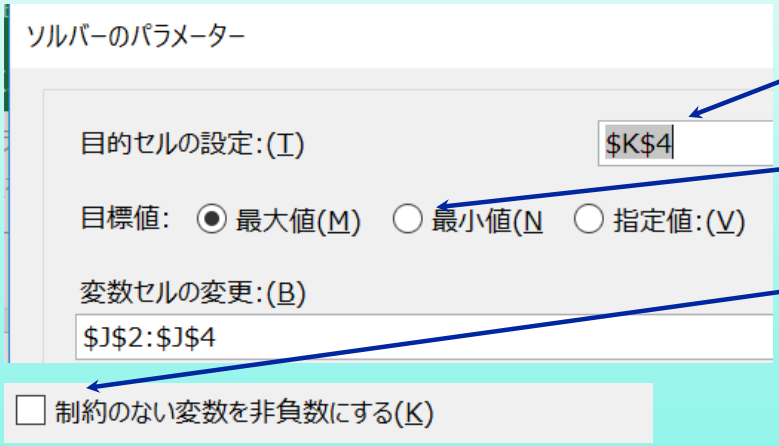
	偏差平方和	
$\hat{\mu}$ 18	327.000	
$\hat{\mu}$ 18.5	324.500	最小値
$\hat{\mu}$ 19	327.000	

標準偏差は，偏差平方和を用いて計算

$$SD = \sqrt{324.5/(10-1)} = 6.005$$

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

5. 正規分布確の率密度の活用

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^ とある標準偏差 σ^ 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^ と σ^ 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

確率密度とは何か

μ =	0	0	0	0	0	0
σ =	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, μ, σ, false)			Norm.dist(x, μ, σ, true)		

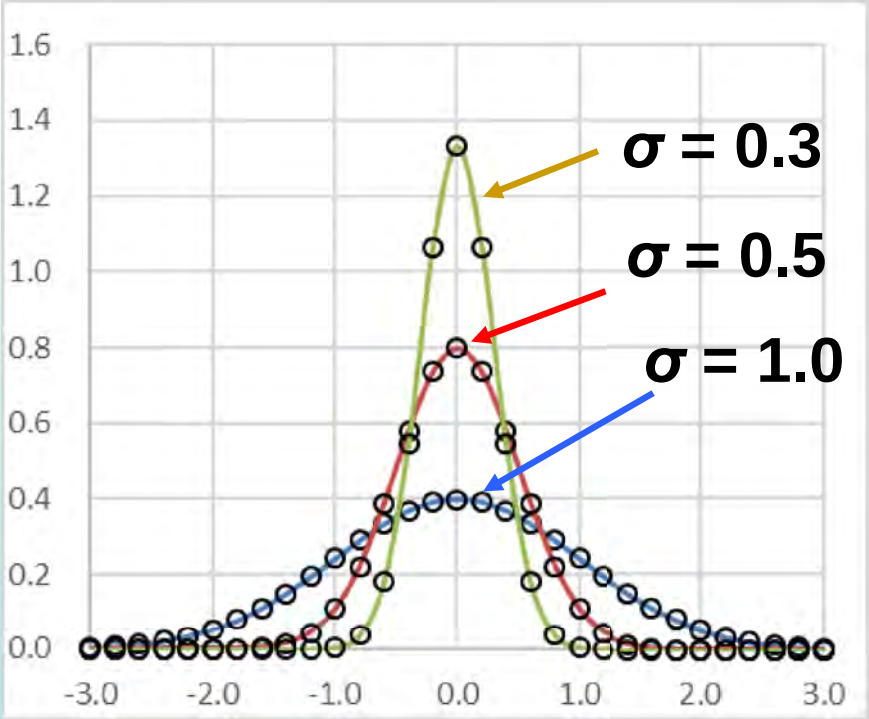
正規分布の確率密度関数： $f(x)$ or $\varphi(x)$

〃　　累積分布関数：$F(x)$ or $\Phi(x)$

（下側確率）

正規分布の確率密度関数

- 面積は 1.0
- 関数の値は，確率ではないので，確率密度と言う.
- ある小さな区間とすれば確率となる.

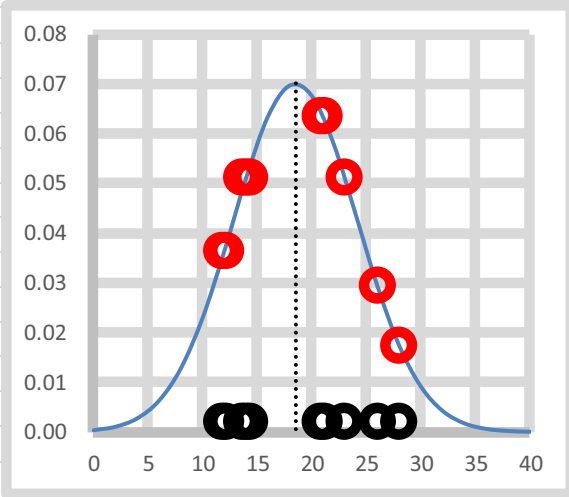


なぜ 積 なのか

- 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 × 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．

高用量群：確率密度の積

		μ^=	18.500	25	18.5
		σ^=	5.696	5.696	3
水平y	拡散x	発毛日y	確率密度	確率密度	確率密度
0.002	11.8	12	0.0365	0.0052	0.0127
0.002	12.2	12	0.0365	0.0052	0.0127
0.002	13.6	14	0.0513	0.0109	0.0432
0.002	14	14	0.0513	0.0109	0.0432
0.002	14.4	14	0.0513	0.0109	0.0432
0.002	20.8	21	0.0636	0.0547	0.0940
0.002	21.2	21	0.0636	0.0547	0.0940
0.002	23	23	0.0513	0.0658	0.0432
0.002	26	26	0.0294	0.0690	0.0058
0.002	28	28	0.0174	0.0610	0.0009
		積	1.91E-14	2.85E-17	2.56E-17
		ln(積)	-31.59	-38.10	-38.20
		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)	



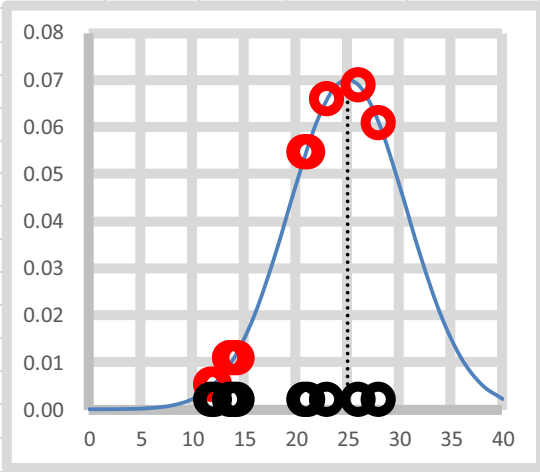
正規分布の確率密度

=NORM.DIST(y_i, μ^, σ^, FALSE)

ソルバーで ln(積)を最大化

平均 μ^ を大きいほうに移動

		μ^=	25.000	25	18.5
		σ^=	5.696	5.696	3
水平y	拡散x	発毛日y	確率密度	確率密度	確率密度
0.002	11.8	12	0.0052	0.0052	0.0127
0.002	12.2	12	0.0052	0.0052	0.0127
0.002	13.6	14	0.0109	0.0109	0.0432
0.002	14	14	0.0109	0.0109	0.0432
0.002	14.4	14	0.0109	0.0109	0.0432
0.002	20.8	21	0.0547	0.0547	0.0940
0.002	21.2	21	0.0547	0.0547	0.0940
0.002	23	23	0.0658	0.0658	0.0432
0.002	26	26	0.0690	0.0690	0.0058
0.002	28	28	0.0610	0.0610	0.0009
		積	2.85E-17	2.85E-17	2.56E-17
		ln(積)	-38.10	-38.10	-38.20
		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)	



正規分布の確率密度

=NORM.DIST(y_i, μ^, σ^, FALSE)

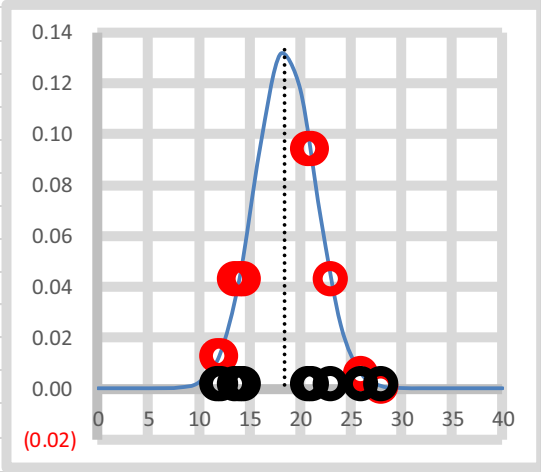
左側の点の確率密度が小さくなる

2017.11.18 高橋行雄

31

標準偏差を小さく

		μ^=	18.500	25	18.5
		σ^=	3.000	5.696	3
水平y	拡散x	発毛日y	確率密度	確率密度	確率密度
0.002	11.8	12	0.0127	0.0052	0.0127
0.002	12.2	12	0.0127	0.0052	0.0127
0.002	13.6	14	0.0432	0.0109	0.0432
0.002	14	14	0.0432	0.0109	0.0432
0.002	14.4	14	0.0432	0.0109	0.0432
0.002	20.8	21	0.0940	0.0547	0.0940
0.002	21.2	21	0.0940	0.0547	0.0940
0.002	23	23	0.0432	0.0658	0.0432
0.002	26	26	0.0058	0.0690	0.0058
0.002	28	28	0.0009	0.0610	0.0009
		積	2.56E-17	2.85E-17	2.56E-17
		ln(積)	-38.20	-38.10	-38.20
		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)	



正規分布の確率密度

=NORM.DIST(y_i, μ^, σ^, FALSE)

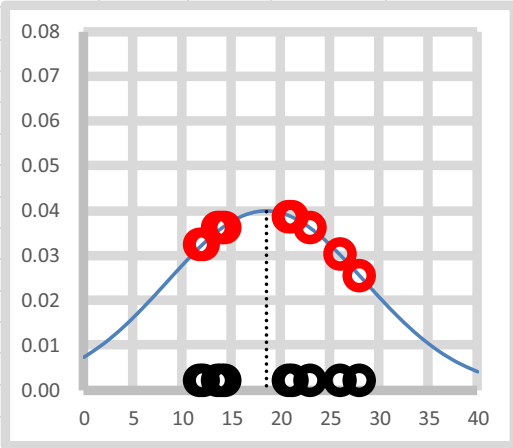
確率密度が小さい両裾に移動

2017.11.18 高橋行雄

32

標準偏差を大きく

		μ^=	18.500	25	18.5
		σ^=	10.000	5.696	3
水平y	拡散x	発毛日y	確率密度	確率密度	確率密度
0.002	11.8	12	0.0323	0.0052	0.0127
0.002	12.2	12	0.0323	0.0052	0.0127
0.002	13.6	14	0.0361	0.0109	0.0432
0.002	14	14	0.0361	0.0109	0.0432
0.002	14.4	14	0.0361	0.0109	0.0432
0.002	20.8	21	0.0387	0.0547	0.0940
0.002	21.2	21	0.0387	0.0547	0.0940
0.002	23	23	0.0361	0.0658	0.0432
0.002	26	26	0.0301	0.0690	0.0058
0.002	28	28	0.0254	0.0610	0.0009
		積	2.02E−15	2.85E−17	2.56E−17
		ln(積)	−33.84	−38.10	−38.20
		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)	



正規分布の確率密度

=NORM.DIST(y_i, μ^, σ^, FALSE)

全体的に確率密度が小さくなる

確率密度の ln(積)の最大化

μ^=		18.500	25	18.5
σ^=		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

σ^ は，SD = √ 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)

6. 正規分布の上側確率の活用

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.

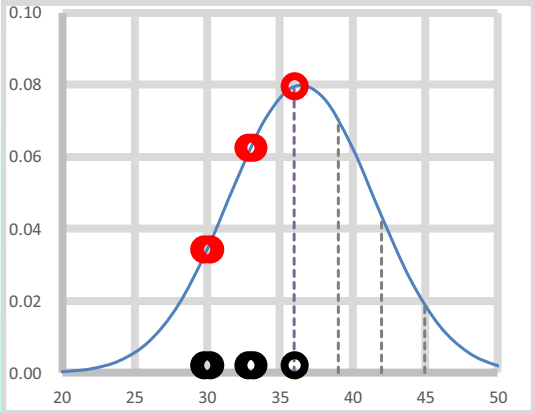
1	2	3	4	5	6	7	8	9	10
30	30	33	33	36	36+	36+	36+	36+	36+

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，発現日が36日以上のどこかなので，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.

対照群：確率密度と上側確率

		μ ^=		36.500	
		σ ^=		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	



μ^=36.5

σ^= 5

上側確率： 1 – Norm.Dist(x_i , μ^, σ^, true)

積の積について自然対数

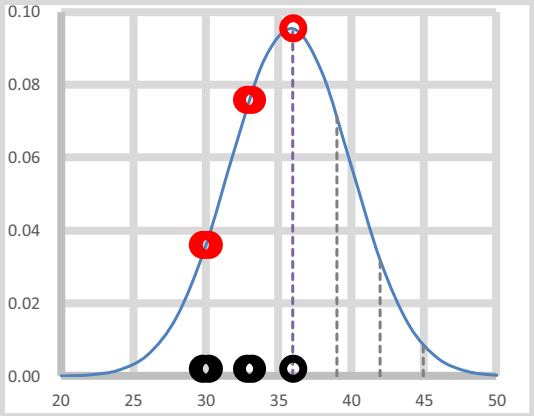
- （確率密度の積）と（上側確率の積）の積を計算し，その自然対数を取る．
- ln（積の積）を最大化するように，ソルバーで μ^=36.5 と σ^=5 を変化させる．

注）定量限界値がある場合には，左側打ち切りとなり，下側確率が使われている．

- 故障・寿命データの場合には，ワイブル分布が使われているが，考え方は同じである．

Ln（積の積）の最大化

		μ ^=		35.835	
		σ ^=		4.172	
水平y	拡散x	発毛日		確率密度	上側確率
0.002	29.8	30		0.0360	
0.002	30.2	30		0.0360	
0.002	32.8	33		0.0759	
0.002	33.2	33		0.0759	
0.002	36	36		0.0956	
0.002		36	+		0.4842
0.002		36	+		0.4842
0.002		36	+		0.4842
0.002		36	+		0.4842
0.002		36	+		0.4842
		積		7.12E−07	2.66E−02
		積の積		1.90E−08	
		ln（積の積）		−17.78	



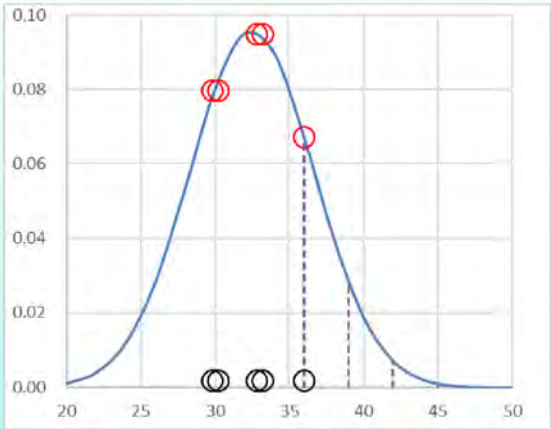
μ^=35.835

σ^= 4.172

ソルバーで，ln（積の積）を最大化するように，μ^ と σ^ を変化させる．

平均値を小さいほうに動かす

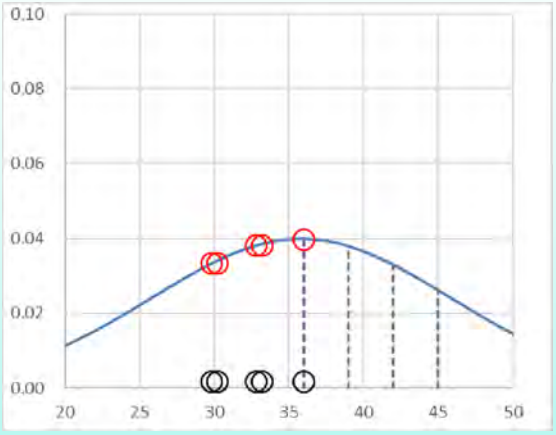
		μ^=		32.500	
		σ^=		4.172	
水平y	拡散x	発毛日		確率密度	上側確率
0.002	29.8	30		0.0799	
0.002	30.2	30		0.0799	
0.002	32.8	33		0.0949	
0.002	33.2	33		0.0949	
0.002	36	36		0.0673	
0.002		36	+		0.2007
0.002		36	+		0.2007
0.002		36	+		0.2007
0.002		36	+		0.2007
0.002		36	+		0.2007
		積		3.87E-06	3.26E-04
		積の積		1.26E-09	
		ln(積の積)		-20.49	



上側確率が小さくなる

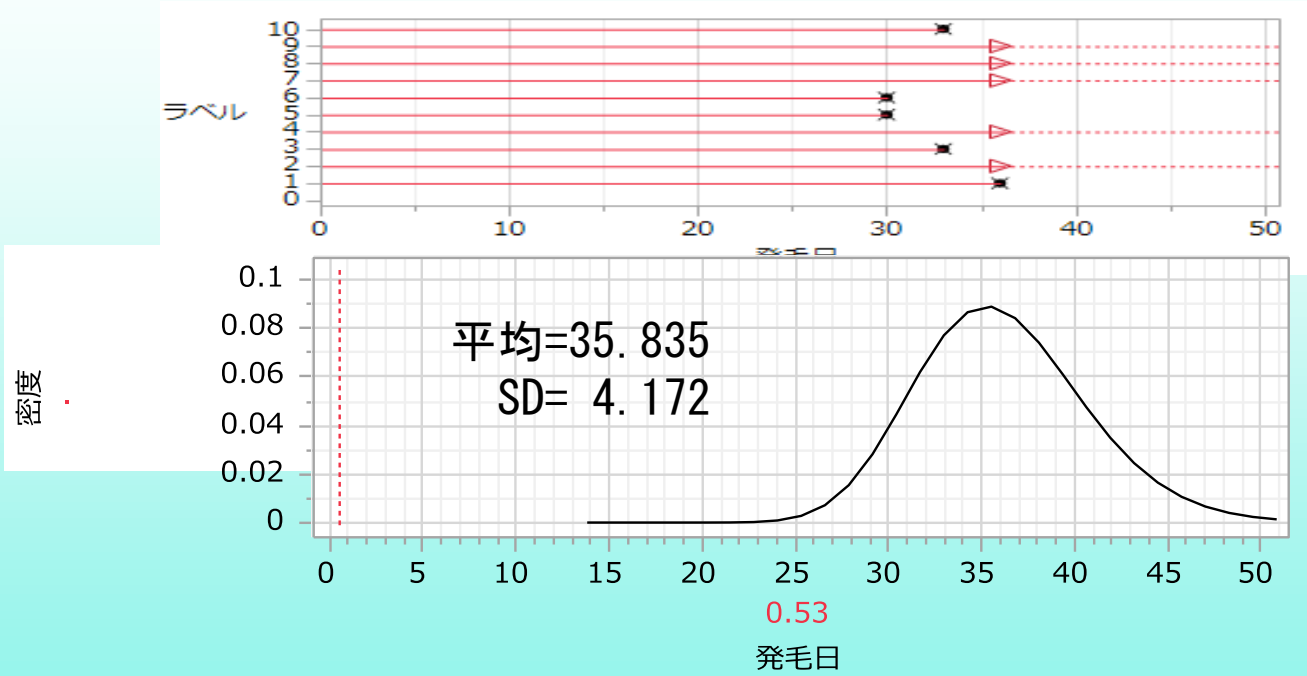
標準偏差を大きくする

		μ^=		35.835	
		σ^=		10.000	
水平y	拡散x	発毛日		確率密度	上側確率
0.002	29.8	30		0.0337	
0.002	30.2	30		0.0337	
0.002	32.8	33		0.0383	
0.002	33.2	33		0.0383	
0.002	36	36		0.0399	
0.002		36	+		0.4934
0.002		36	+		0.4934
0.002		36	+		0.4934
0.002		36	+		0.4934
0.002		36	+		0.4934
		積		6.63E-08	2.92E-02
		積の積		1.94E-09	
		ln(積の積)		-20.06	



確率密度が全体的に小さくなる

JMPの結果（再掲）



ln（積の積）が最大であることの確認

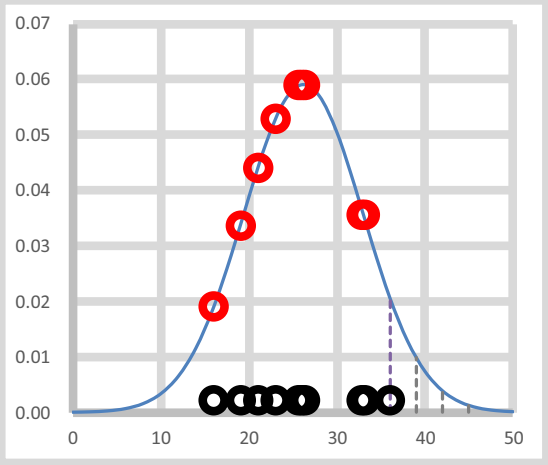
μ^=	35.835		35.835		36.500	
σ^=	4.172		5.000		4.172	
発毛日	確率密度	上側確率	確率密度	上側確率	確率密度	上側確率
30	0.0360		0.0404		0.0284	
30	0.0360		0.0404		0.0284	
33	0.0759		0.0679		0.0673	
33	0.0759		0.0679		0.0673	
36	0.0956		0.0797		0.0949	
36 +		0.4842		0.4868		0.5477
36 +		0.4842		0.4868		0.5477
36 +		0.4842		0.4868		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

中用量群：打ち切りが1匹

		μ^=		26.202	
		σ^=		6.761	
水平y	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)



μ^=26.202

σ^= 6.761

算術平均は，25.5 と過少評価となる

7. 打ち切りを含む回帰分析

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），．．．，（x_n, y_n）とする.
- y_i^ = β_0^ + β_1^ x_i
- 偏差平方和: $SS = \Sigma\,(y_i - y_i\hat{})^2$ を最小化するような（β_0^, β_1^）を求める.
- なお, Excel の Linest 関数でも簡単に計算できる.

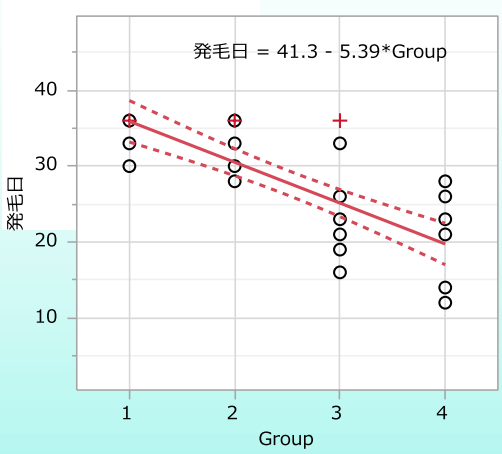
JMPの二変量の関係

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	p値(Prob>F)
全体(修正済み)	39	2441.7750		<.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.

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 y_i^ を平均とみなした正規分布を考える.

Excel での対応 1

- データを（x_1, y_1, δ_1），. . . . ，（x_n, y_n, δ_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 β_0^, β_1^, σ^ に適当な値を設定する.
- 回帰の推定値の計算: y_i^ = β_0^ + β_1^ x_i
- 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.

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 , y_i^, σ^, false)
  - $\delta_i = 1$: 1 - Norm.dist(y_i , y_i^, σ^, 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.

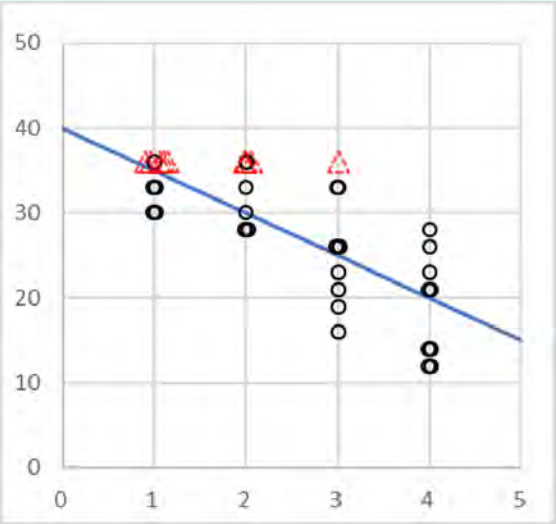
Excel による回帰分析のシート

					β_0 ^=	44.654			
					β_1 ^=	−6.320			対数の和
					σ ^ =	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

打ち切りを考慮した回帰分析

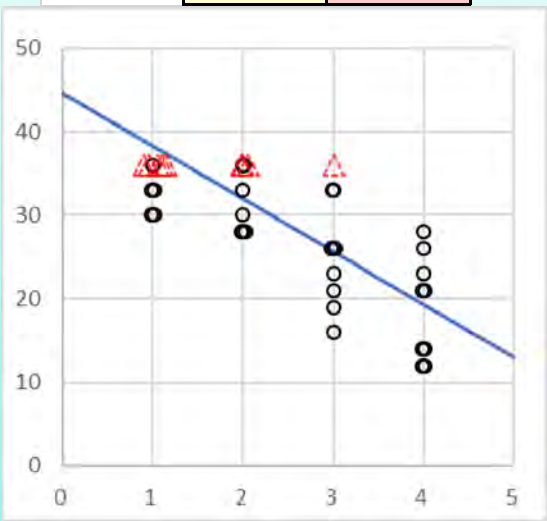
初期値

β_0 ^=	40.000	
β_1 ^=	−5.000	対数の和
σ^ =	5.000	−108.811

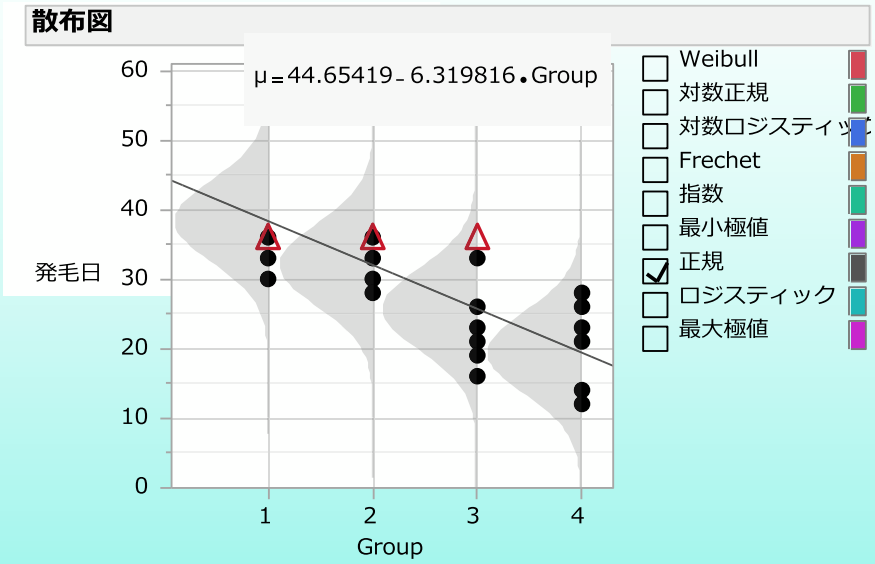


対数の和を最大化

β_0 ^=	44.654	
β_1 ^=	−6.320	対数の和
σ^ =	5.889	−105.845



JMP「寿命の二変量」散布図



打ち切りデータを含む回帰分析．誤差分布に正規分布を設定．△が打ち切りデータとして考慮されている．

- 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．

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%
β0	44.65419	2.5439852	39.66807	49.64031
β1	-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.

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
☐	効果なし	245.7339	250.0583	253.1117	2
	回帰	211.69	218.3567	222.7567	3
☑	別々の位置	209.7529	221.5176	228.1973	5
☑	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（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

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

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

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.

分散共分散行列

共分散行列			
パラメータ	β0	β1	σ
β0	6.4719	-2.1001	0.3662
β1	-2.1001	0.8003	-0.0986
σ	0.3662	-0.0986	0.5938

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

$\hat{y}$の分散

$$\mathrm{Var}(\hat{y}) = \mathrm{Var}(\hat{\beta}_0 + \hat{\beta}_1 x)$$
$$= \mathrm{Var}(\hat{\beta}_0) + 2\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1)\, x_i + \mathrm{Var}(\hat{\beta}_1)\, x_i^2$$

$\boldsymbol{x} = [\,1,\ x\,]$ $\quad \boldsymbol{\Sigma} =$

パラメータ	β0	β1
β0	6.4719	-2.1001
β1	-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．

$$\mathrm{Var}(\hat{y}) = \boldsymbol{x}\,\boldsymbol{\Sigma}\,\mathbf{x}^T$$

Var（y^）および95%信頼区間

β_0 ^=	44.654
β_1 ^=	−6.320
σ^ =	5.889

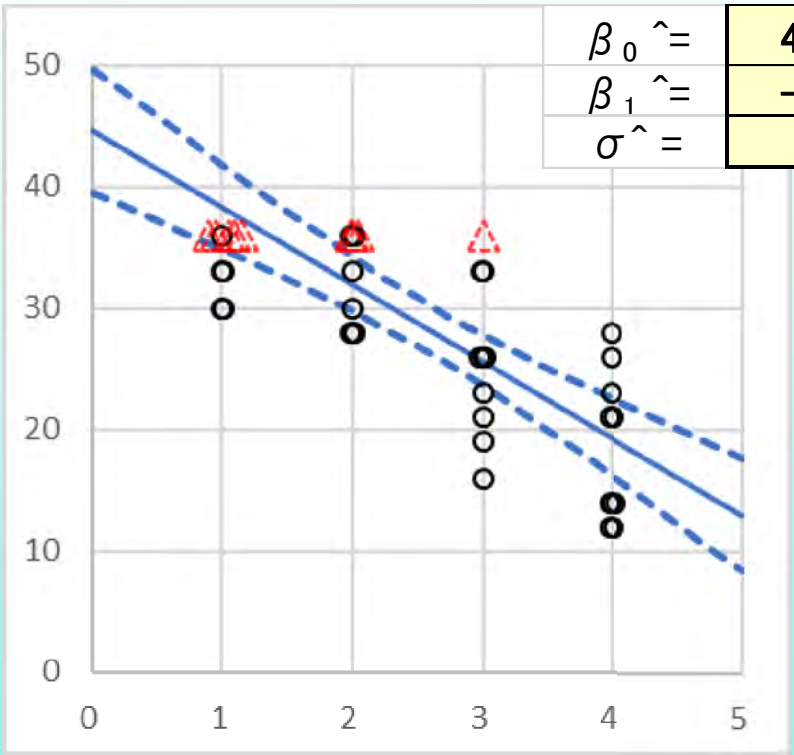
パラメータ	β0	β1	σ
β0	6.471861	−2.10006	0.366203
β1	−2.10006	0.800287	−0.09864
σ	0.366202	−0.09864	0.593759

1	x	y^	Var(y^)	L95%	U95%
1	0.00	44.654	6.472	39.67	49.64
1	0.25	43.074	5.472	38.49	47.66
1	0.50	41.494	4.572	37.30	45.69
1	2.00	32.015	1.273	29.80	34.23
1	2.25	30.435	1.073	28.40	32.46
1	2.50	28.855	0.973	26.92	30.79
1	2.75	27.275	0.974	25.34	29.21
1	3.00	25.695	1.074	23.66	27.73

Excel の計算式 Var(y^)： = Mmult（Mmurt(***x***, Σ)，Transpose(***x***))

95%CI ： = y^±1.96*sqrt(Var(y^))

回帰直線および95%信頼区間

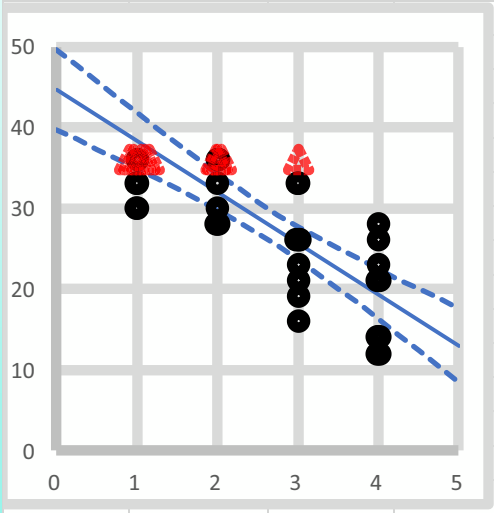


β_0 ^=	44.654
β_1 ^=	−6.320
σ^ =	5.889

Excel シートの一部

回帰直線の座標			
x	y	β0 ^=	44.654
0	44.654	β1 ^=	−6.320
5	13.055	σ^ =	5.889

パラメータ	β0	β1	σ
β0	6.471861	−2.10006	0.366203
β1	−2.10006	0.800287	−0.09864
σ	0.366202	−0.09864	0.593759



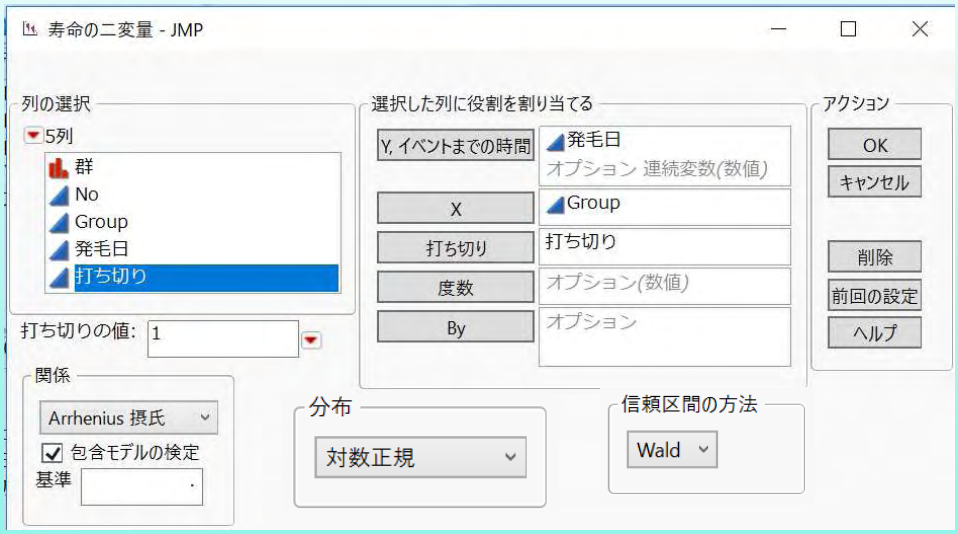
1	x	y^	Var(y^)	L95%	U95%
1	0.00	44.654	6.472	39.67	49.64
1	0.25	43.074	5.472	38.49	47.66
1	0.50	41.494	4.572	37.30	45.69
1	0.75	39.914	3.772	36.11	43.72
1	1.00	38.334	3.072	34.90	41.77
1	1.25	36.754	2.472	33.67	39.84
1	1.50	35.174	1.972	32.42	37.93
1	1.75	33.594	1.573	31.14	36.05
1	2.00	32.015	1.273	29.80	34.23
1	2.25	30.435	1.073	28.40	32.46
1	2.50	28.855	0.973	26.92	30.79
1	2.75	27.275	0.974	25.34	29.21
1	3.00	25.695	1.074	23.66	27.73
1	3.25	24.115	1.275	21.90	26.33

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．

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

第 2 期　医 薬 安 全 性 研 究 会

会員の方は http://biostat.jp/ からこの定例会資料を印刷することができます. PDF の両面印刷でご利用ください. 発表時のスライドは定例会後に PDF 化して閲覧できるようにします. 発表資料中に提示されているデータは csv 形式で提供します. 非会員の方は，会員への登録後に閲覧可能となります. 定例会資料集の著作権は，第 2 期 医薬安全性研究会にあります.

連絡先： toiawase@biostat.jp （全て半角小文字です）
ホームページ： http://biostat.jp/