

# カブトガニのサテライト数の観察データ に関する探索的ポアソン回帰

BioStat研究所(株)  
高橋 行雄

# 一般化線形モデル

---

- ◆ SASのGENMODにより一般化線形モデルが使えるようになった時に、最初に注目したのは、従来からあるLOGISTICプロシジャとの使い分けであった.
- ◆ 最近、植物の種子数を主体した「一般化線形モデル(GLM)ーポアソン回帰ー」が注目され、正規分布を前提とした回帰分析とは異なる切り口が提唱されている.

# カウンタデータの事例

---

- ◆ アグレスティ著，渡邊ら訳（2003），カテゴリーカルデー解析入門，サイエンティスト社，110-127, 168-179.
  - 雌のカブトガニに連結する雄のサテライト数（Satellite数）の例を取り上げる.
  - 173匹のカブトガニ：説明変数として甲羅の色，後体部の棘の状態，甲羅の幅，体重の4変数，応答変数としてサテライト数

## 雌のカブトガニに連結する雄のサテライト数

col or	spi ne	width	weight	sat ell	col or	spi ne	width	weight	sat ell	col or	spi ne	width	weight	sat ell	col or	spi ne	width	weight	sat ell
2	3	28.3	3.050	8	3	1	28.5	3.250	9	4	3	23.5	1.900	0	2	1	28.0	2.900	4
3	3	22.5	1.550	0	3	3	28.9	2.800	4	2	2	24.0	1.700	0	4	3	25.8	2.250	10
1	1	26.0	2.300	9	2	3	28.2	2.600	6	2	1	29.7	3.850	5	2	3	27.9	3.050	7
3	3	24.8	2.100	0	2	3	25.0	2.100	4	2	1	26.8	2.550	0	2	3	24.9	2.200	0
3	3	26.0	2.600	4	2	3	28.5	3.000	3	4	3	26.7	2.450	0	2	1	28.4	3.100	5
2	3	27.9	2.800	6	3	1	25.9	2.550	4	4	3	22.5	1.475	4	4	3	27.0	2.625	0
3	3	27.5	3.100	3	2	3	25.8	2.300	0	2	3	26.2	2.025	2	2	2	24.5	2.000	0
1	1	26.1	2.800	5	4	3	27.0	2.250	3	2	1	24.9	2.300	6					
1	1	27.7	2.500	6	2	3	28.5	3.050	0	1	2	24.5	1.950	6					
2	1	30.0	3.300	5	4	1	25.5	2.750	0	2	3	25.1	1.800	0					

注釈: color=色 (1=やや明るい, 2=中くらい, 3=やや暗い, 4=暗い);

Spine=後体部の棘の状態 (1=いずれも正常, 2=一方が摩耗または破損している, 3=いずれも摩耗または破損している);

width=甲羅の幅(cm); weight=重さ(kg); satell=サテライト数.

出典: <http://lib.stat.cmu.edu/datasets/agresti>. 2019年7月24日 アクセス.

# 文献での探索的な解析

---

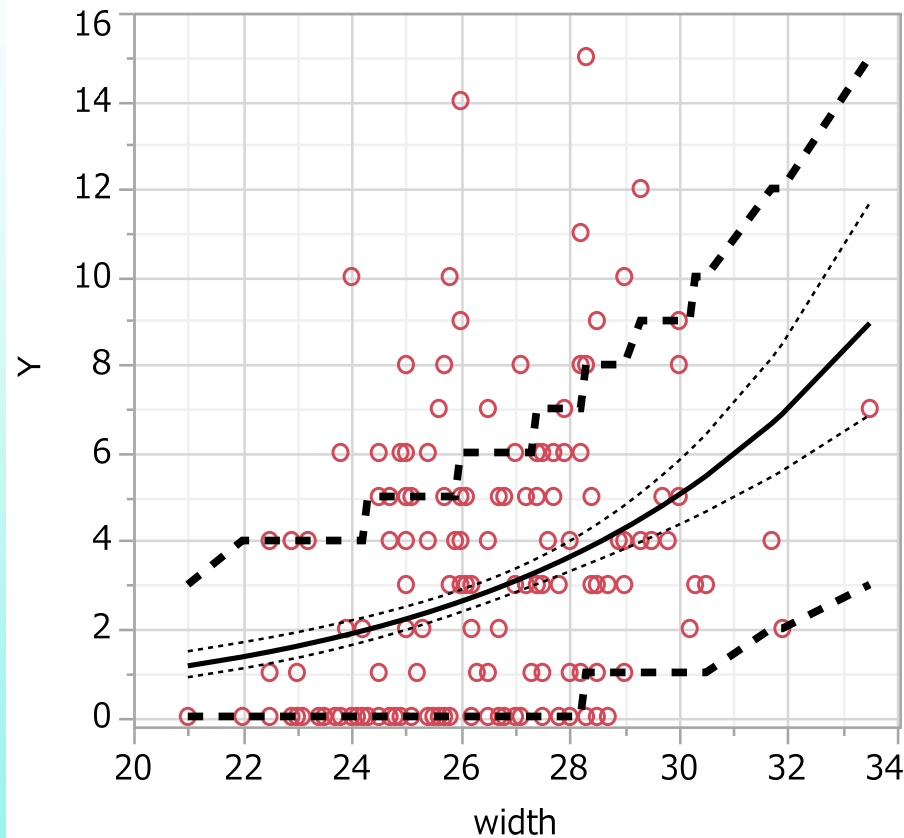
- ◆ 甲羅の幅をX軸，サテライト数をY軸とした散布図と共に，対数リンクによるポアソン回帰
  - その後の解析では甲羅の幅を8区分とし区分内のカブトガニの数とサテライト数の合計を算出し，カブトガニの数をオフセットとした解析を主体
  - 探索的な解析では，サテライト数が(0, 1以上)の2値データとして，ロジスティック回帰を主体にした解析方法が提示

# ポアソン分布のあてはめ

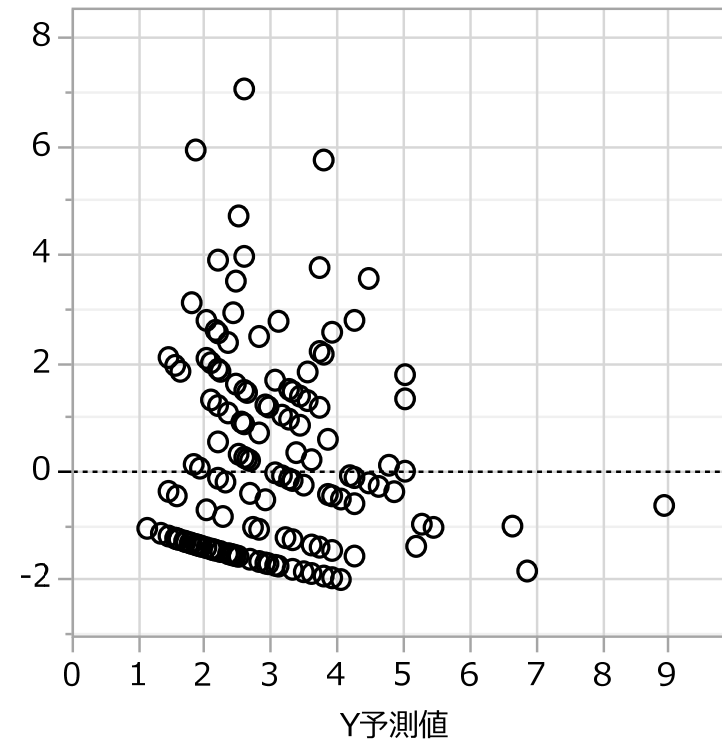


平均が異なる部分集団が存在するのか？

# 甲羅の幅についての回帰



測値 vs. Pearson残差



## 対数リンクのポアソン回帰でも過分散

# 過分散の調整

- ◆ 適合度のピアソンのカイ2乗統計量を自由度で除した調整パラメータ  $\phi$  を使う.

適合度統計量	カイ2乗	自由度	p値(Prob>ChiSq)
Pearson	544.1570	171	<.0001*
デビアンس	567.8786	171	<.0001*

$$\phi = 544.1570 / 171 = 3.1822$$

- ◆ ポアソン回帰で得られた分散を $\phi$  倍する.



# 過分散パラメータの適用

## 調整前

パラメータ推定値				
項	推定値	標準誤差	尤度比カイ2乗	p値
切片	-3.3048	0.5422	36.8670	<.0001*
甲羅の幅	0.1640	0.0200	64.9131	<.0001*

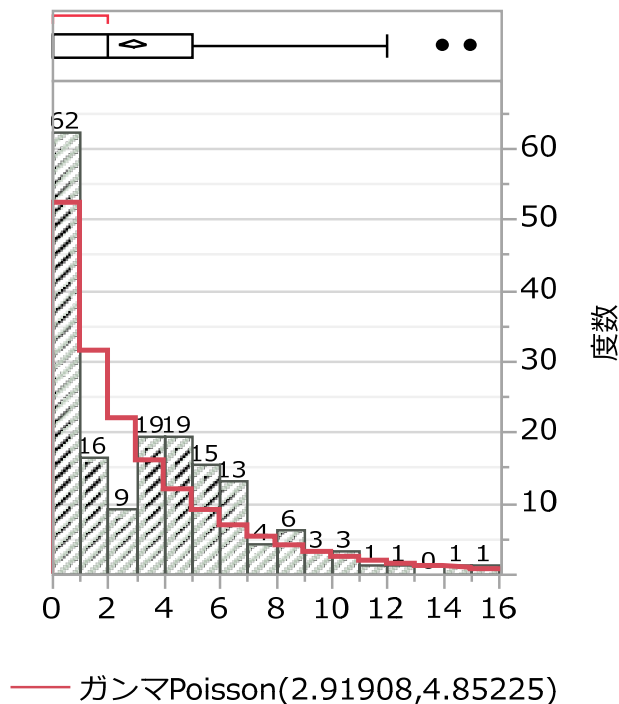
## 調整後

項	推定値	標準誤差	尤度比カイ2乗	p値
切片	-3.3048	0.9673	11.5854	0.0007*
甲羅の幅	0.1640	0.0356	20.3988	<.0001*

$$SE' = \sqrt{\phi SE^2} = \sqrt{3.1822 \times 0.0200^2} = 0.0356$$

# ガンマ・ポアソン分布のあてはめ

## サテライト数



## ガンマPoissonのあてはめ

### パラメータ推定値

種類	パラメータ	推定値	下側95%CL	上側95%CL
位置	$\lambda$	2.9191	2.4136	3.5571
過分散	$\sigma$	4.8522	3.6559	6.6859

### 指標

(-2)*対数尤度	767.4092
AICc	771.4798
BIC	777.7158

### 適合度検定

Pearsonのカイ2乗

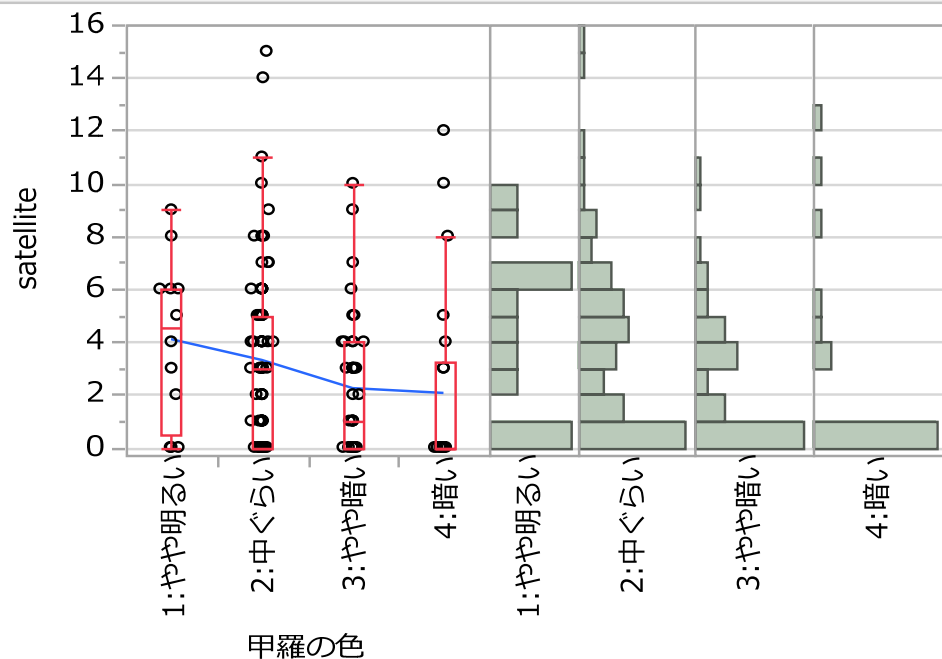
X2	Prob>X2
120.3656	0.9990

注:  $H_0$  = ガンマPoisson分布からのデータ。 p値が小さい場合は $H_0$ を棄却。

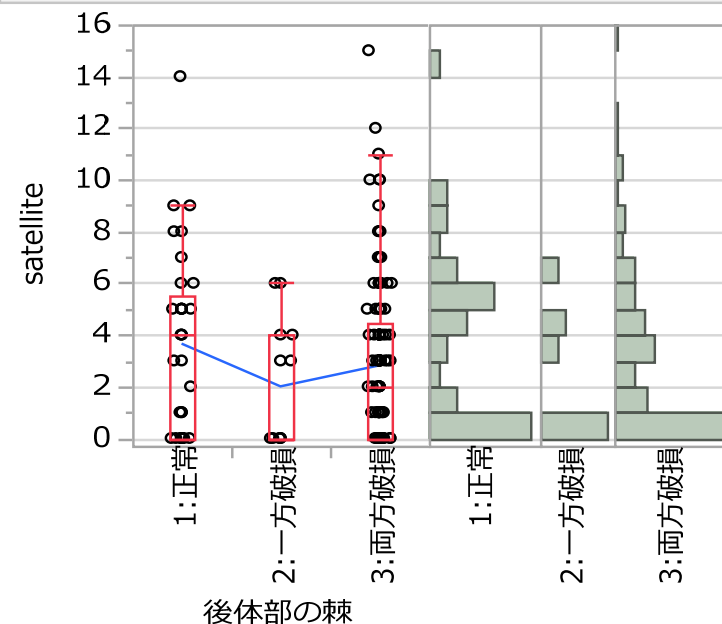
## あてはまりは棄却できない. OKなのか？

# 過分散の解消のための層別解析

甲羅の色によるsatelliteの一元配置分析



後体部の棘によるsatelliteの一元配置分析



◆ 甲羅の色が暗くなるとサテライト数が減少

# 層別解析の結果

---

- ◆ 雌の甲羅の色
  - 暗くなるに従いゼロ・カウントが増える
  - " サテライト数の平均値が減少傾向
- ◆ 雌の後体部の棘の状態
  - 正常の場合には, サテライト数の5匹に山が
  - サテライト数の平均値は同程度
- ◆ 過分散の解消とはならない

# 2変数の組合せ

甲羅の色	棘の状態	N	平均	分散	分散/平均
1:やや明るい	1:正常	9	4.44	10.53	2.37
	2:一方破損	2	4.50	4.50	1.00
	3:両方破損	1	0.00	-	-
2:中ぐらい	1:正常	24	3.29	12.13	3.68
	2:一方破損	8	1.75	6.21	3.55
	3:両方破損	63	3.49	10.03	2.87
3:やや暗い	1:正常	3	5.33	10.33	1.94
	2:一方破損	4	1.75	4.25	2.43
	3:両方破損	37	2.03	6.25	3.08
4:暗い	1:正常	1	0.00	-	-
	2:一方破損	1	0.00	-	-
	3:両方破損	20	2.25	13.99	6.22
	全体	173	2.92	9.91	3.40

◆ 過分散は解消しない。

◆ 甲羅の色と棘の状態に関連がある。

# 甲羅の幅と体重

- ◆ サテライト数に対する2変数の影響を対数リンクによるポアソン2変量回帰

項	推定値	標準誤差	尤度比カイ2乗	p値
切片	-1.2952	0.8989	2.0691	0.1503
甲羅の幅	0.0461	0.0467	0.9658	0.3257
体重	0.4470	0.1586	7.9780	0.0047

- ◆ 甲羅の幅の推定値は, **0.0461**
- ◆ 体重の推定値は, **0.4470**
- ◆ 尤度比検定の結果は, **体重のみが有意**  
**過分散の調整なし**

# 内部構造の可視化

---

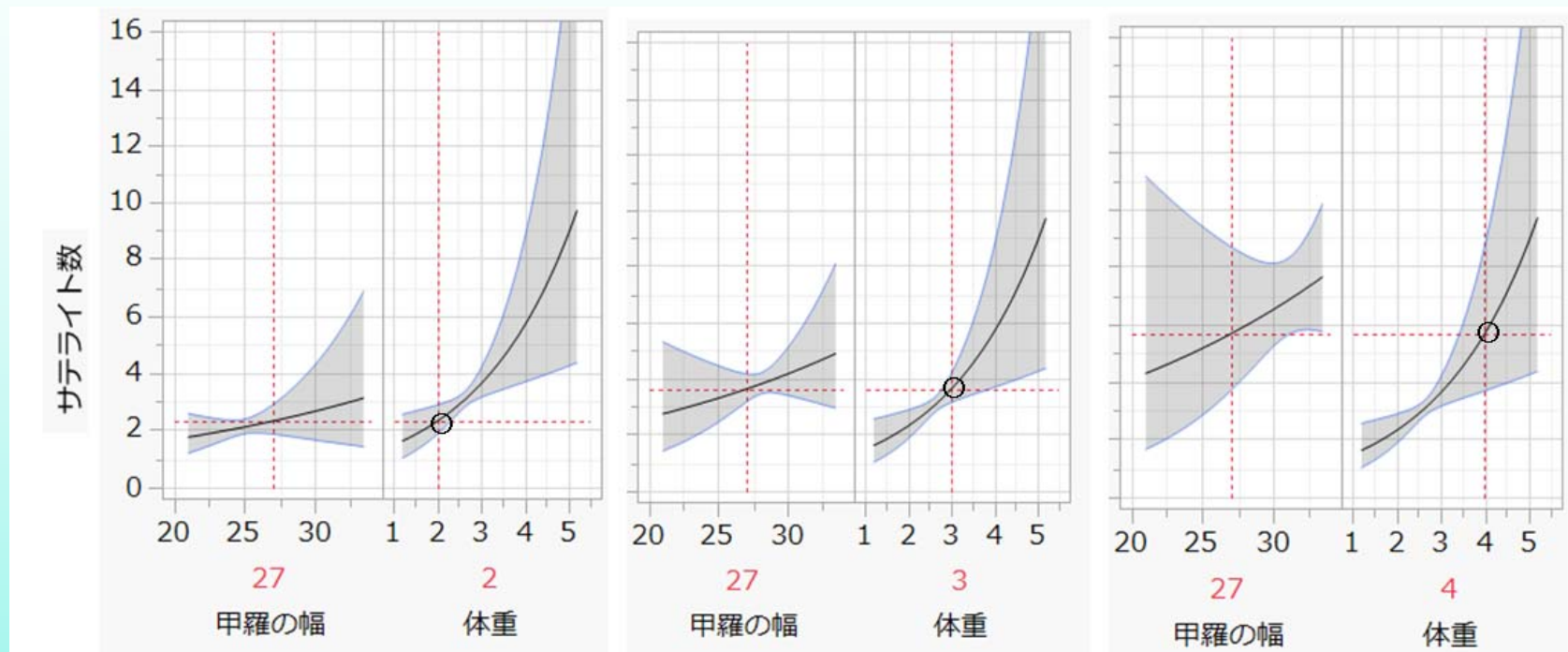
## ◆ 2変量回帰での係数の解釈

- 単独では共に関連があるが, 2変量で他方が関連なし, あるいは, 負の場合の解釈.

## ◆ JMPの「予測プロファイル」で内部構造の可視化する.

- 体重を(2, 3, 4 kg)と変化させた場合の甲羅の幅がサテライト数に及ぼす影響を図示
- 甲羅の幅は, 体重を固定したときにサテライト数に影響を与えていない.

# 甲羅の幅と体重の影響

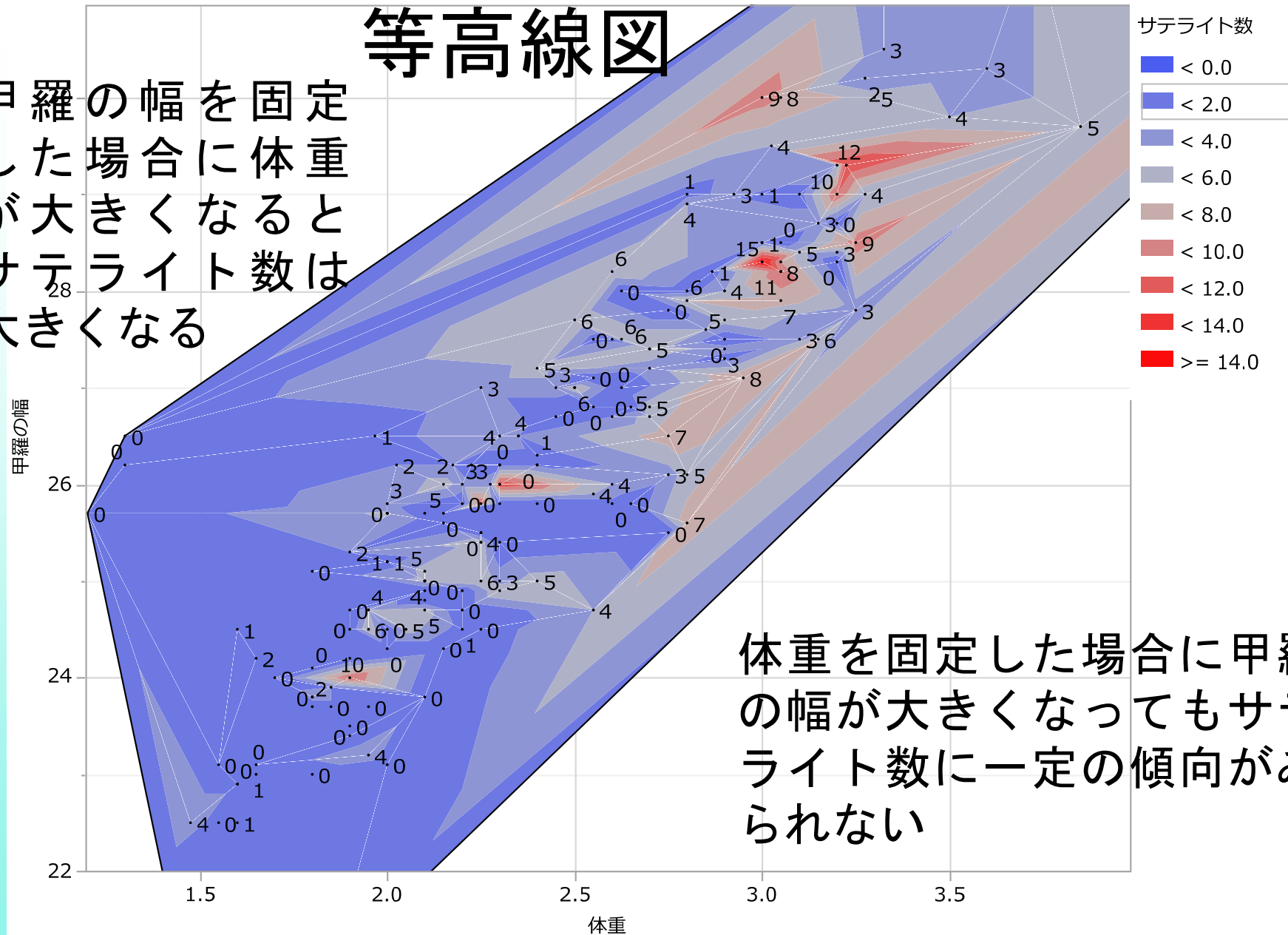


体重と甲羅の幅の2変数のポアソン回帰の結果，体重のみがサテライト数に影響している。



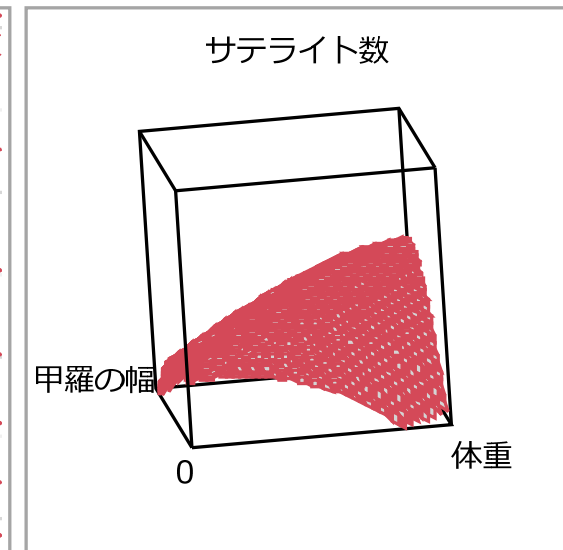
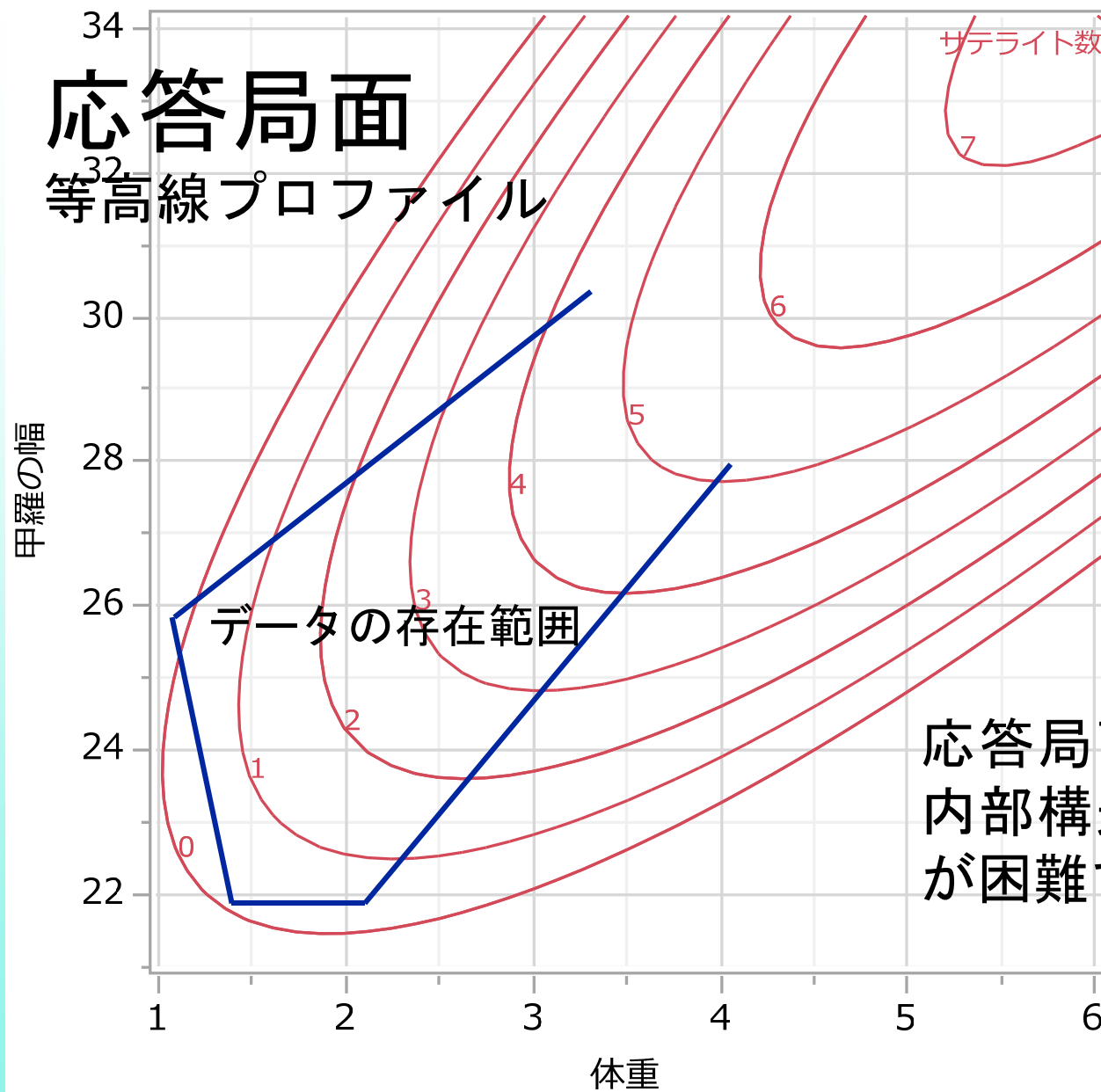
# 等高線図

甲羅の幅を固定した場合に体重が大きくなるとサテライト数は大きくなる



体重を固定した場合に甲羅の幅が大きくなってもサテライト数に一定の傾向がみられない

# 応答局面 等高線プロファイル



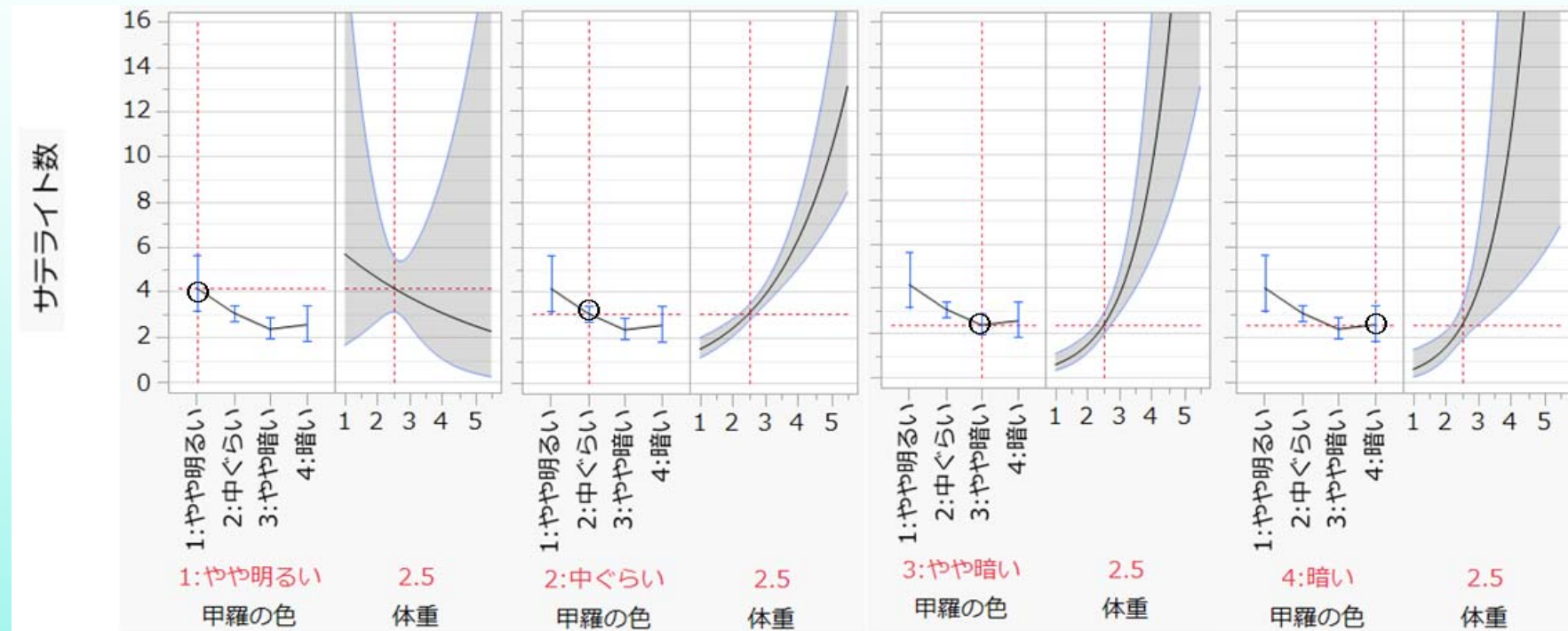
応答局面解析では、複雑な内部構造を可視化することが困難である。

# 甲羅の色と体重

---

- ◆ 甲羅の色を対比型のダミー変数とする
- ◆ 交互作用モデル
  - 甲羅の色, 体重, (甲羅の色 × 体重)
  - 回帰パラメータによる解釈は難解
- ◆ 「予測プロファイル」の機能
  - 甲羅の色ごとの体重の増加によるサテライト数との関連を概観

# 甲羅の色と体重でのポアソン回帰

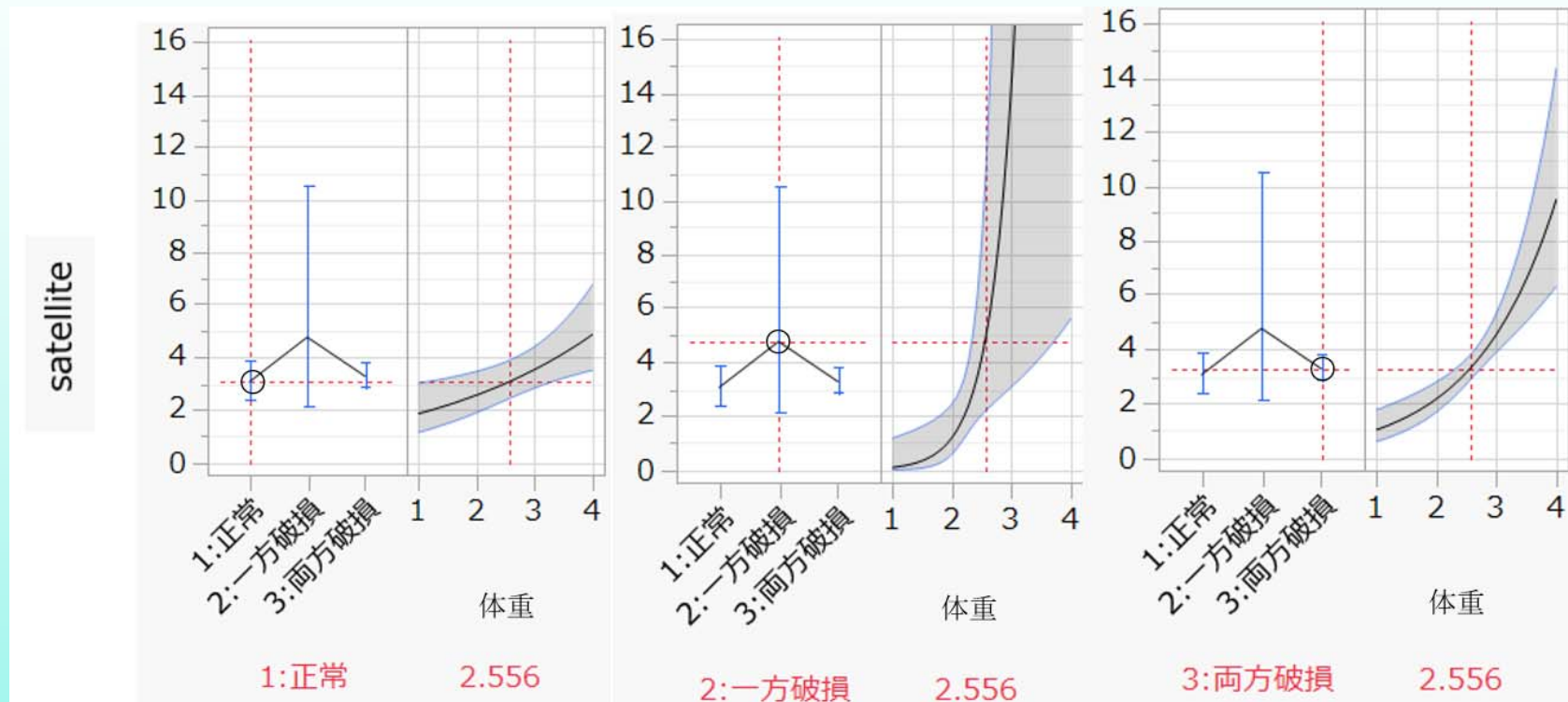


甲羅の色がやや明るい場合にサテライト数は、体重により変化がない。交互作用が示唆

# 後退部の棘の状態を含めた解析

- ◆ 甲羅の色が「**中くらい**」の場合
  - 後体部の棘：(正常：両方破損) = (24：63)
- ◆ サテライト数との関連を甲羅の色が「中くらい」に限定して関連を調べる
  - 後体部の棘が「正常」の場合に体重が増えればサテライト数は微増.
  - 「一方破損」および「両方破損」では、体重が増えた場合にサテライト数が急増

# 棘の状態と体重でのポアソン回帰



棘が正常の場合サテライト数に大して  
体重の影響は軽微

# 交互作用がある場合の解析

---

- ◆ 交互作用が疑われるような探索的な解析を行うためには、各種のグラフ表示が欠かせない.
- ◆ JMPの多彩なグラフ表示を活用し、カブトガニの各種の変数とサテライト数の関連を浮き彫りにしてきたが、今ひとつである.
- ◆ 全体を俯瞰するためにJMPの「グラフ・ビルダー」が有益である.

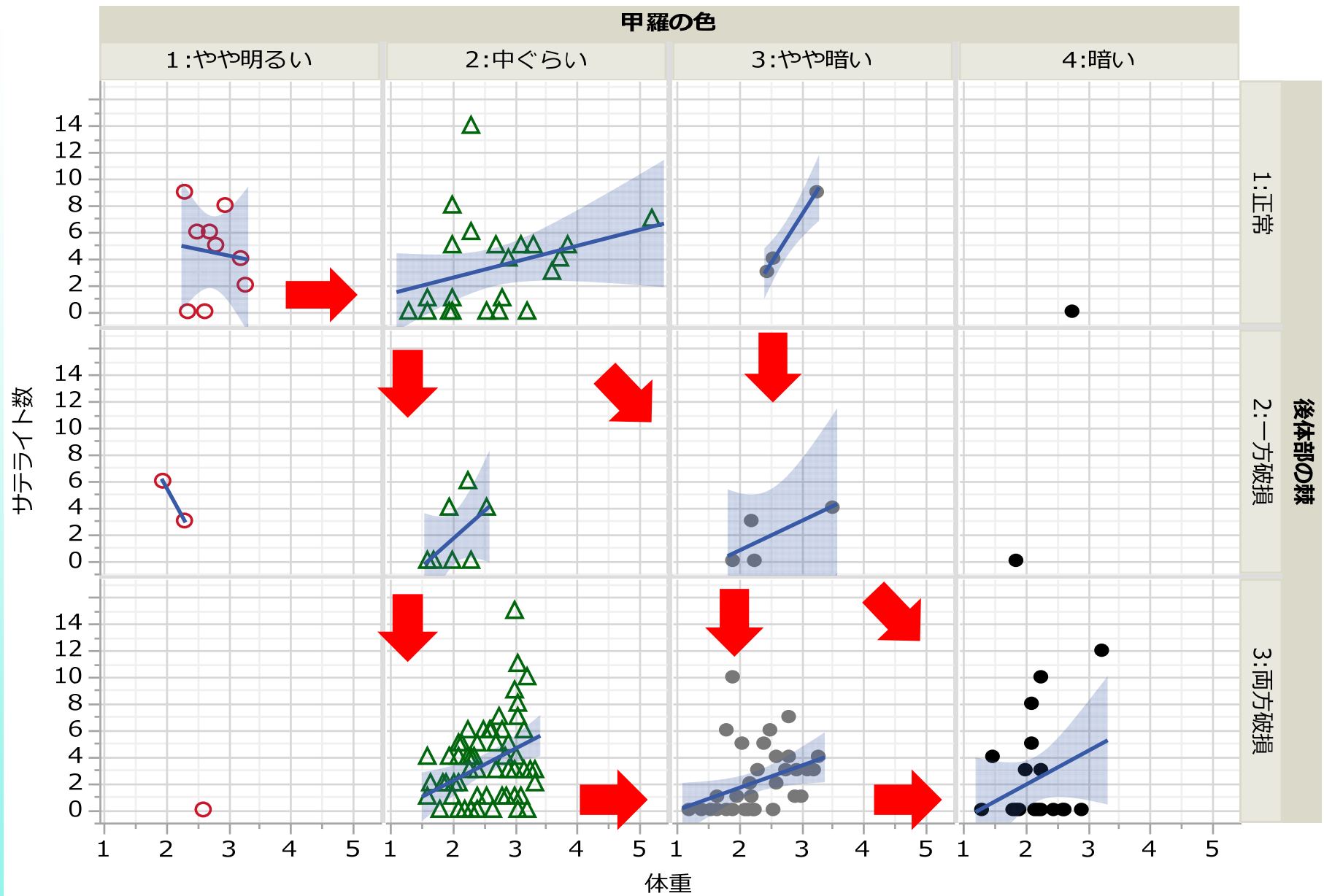
# JMPのグラフ・ビルダー

---

- ◆ Sプラスに同様の機能があり愛用していた
- ◆ JMPグラフ・ビルダーは、Sプラスの機能を大幅に凌駕する探索的な統計解析を支援するツールとして優れている。
- ◆ 交互作用がある場合には、図を作成して解釈することが鉄則である。
  - 交互作用を把握するための作図は容易ではないが、グラフ・ビルダーで容易に把握できる



X:体重,Y:サテライト数



2019.09.06 高橋行雄

# 交互作用の解釈

---

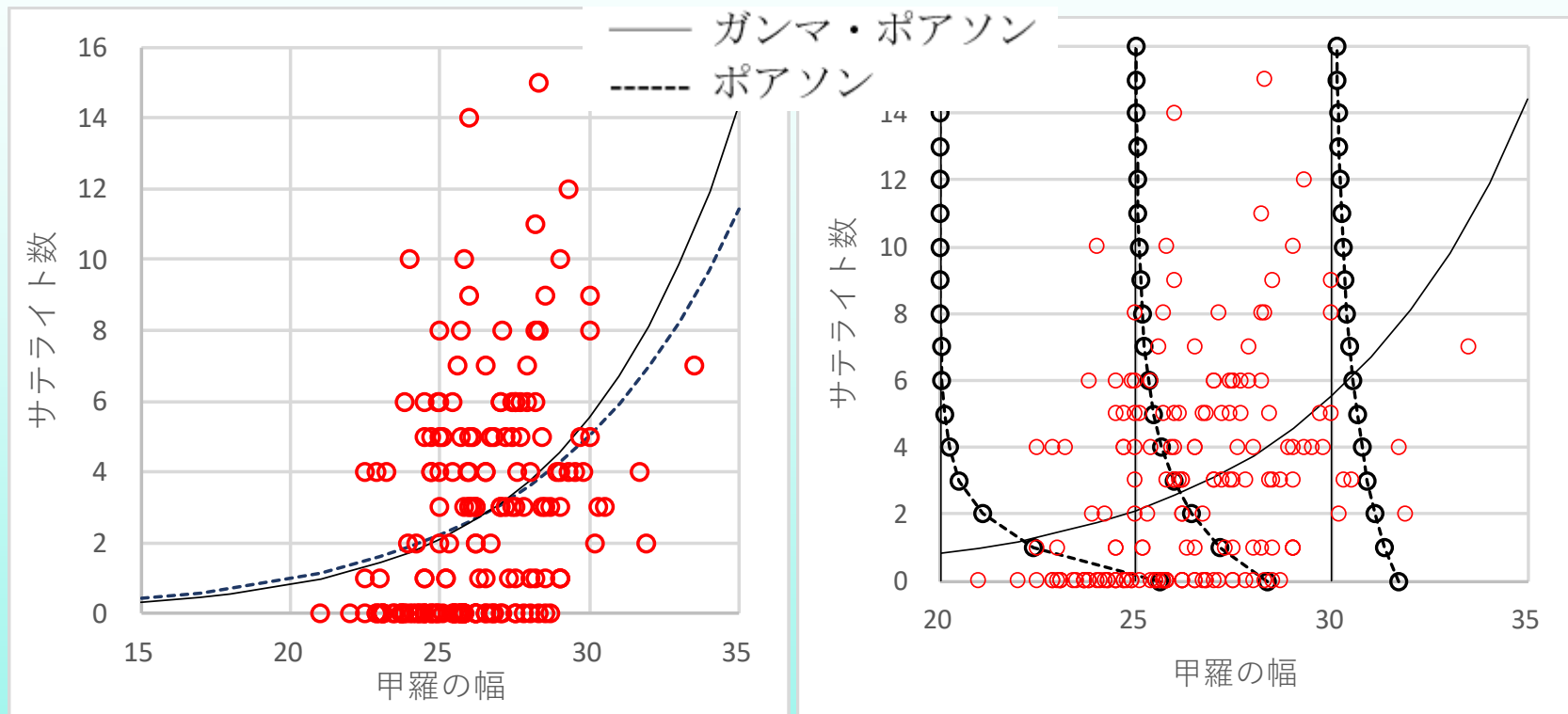
- ◆ 甲羅の色が暗くなるにつれて後体部の棘の破損が進む
- ◆ それに伴い、体重の軽い雌ほど連結する雄のサテライト数が減少
- ◆ 甲羅の色が暗くなり、後体部の棘の状態が悪くなる“加齢現象”により、体重の軽い雌ほど連結する雄のサテライト数が減少する
- ◆ そのため、ゼロ・カウントが多い過分散となったと推測される。

# グラフ・ビルダー再考

---

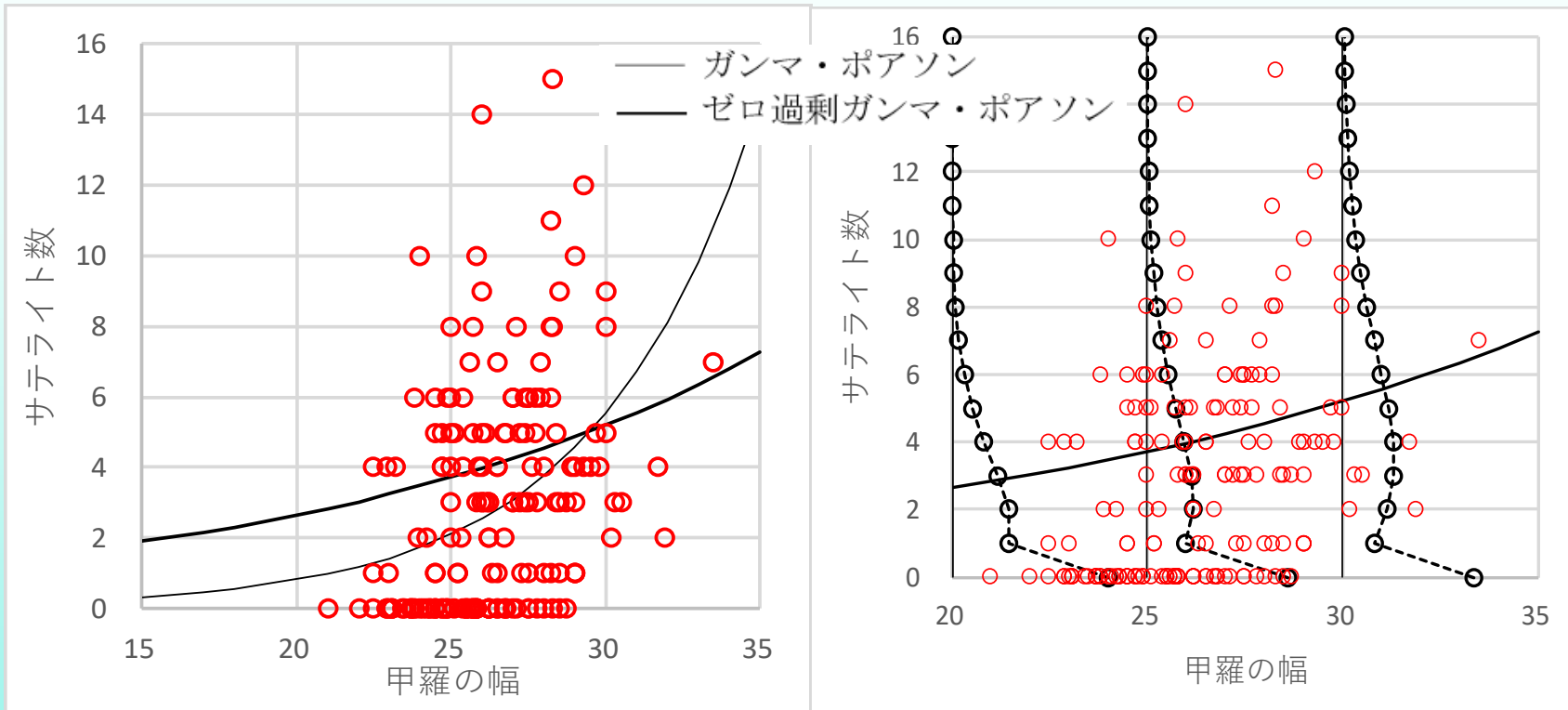
- ◆ 2つの順序変数を組み合わせて $(3 \times 4)$ のセル別に体重とサテライト数の散布図に回帰直線と95%信頼区間が描けることにより、変数間の関係が一覧でき、内部構造の可視化に有益であった。
- ◆ ただし、ポアソン回帰ではなく、 $(3 \times 4)$ のセル別の通常の回帰の結果となっているので、解釈には注意が必要である。

# ガンマ・ポアソン回帰



甲羅の幅が大きい時にガンマ・ポアソン分布の仮定は否定的

# ゼロ過剰ガンマ・ポアソン回帰



甲羅の幅が大きい時にゼロ過剰ガンマ・ポアソン分布の仮定は否定的

# まとめ

---

- ◆ 探索的ポアソン回帰にJMPのプロファイル機能が、データの内部構造を可視化するために有効であった.
- ◆ 過分散の場合にポアソン回帰の結果は誇張されすぎるので、過分散の調整は必須である.
- ◆ ガンマ・ポアソン回帰、ゼロ過剰ガンマ・ポアソン回帰などの適用には注意が必要である.