

KUDH Basics

統計ソフトウェア「R」ワークショップ

4限目

一般化線形モデル

講師：山崎 大暉

(立命館大学・日本学術振興会)

今日の流れと到達目標

1. Rのlm()関数で線形モデルを記述しよう

- 関数に渡すデータフレームを用意する
- モデルの記法を知る
- ダミー変数を用いた質的変数の表現を理解する
- 正規線形モデルを使って分析する

2. glm()関数で一般化線形モデルを使おう

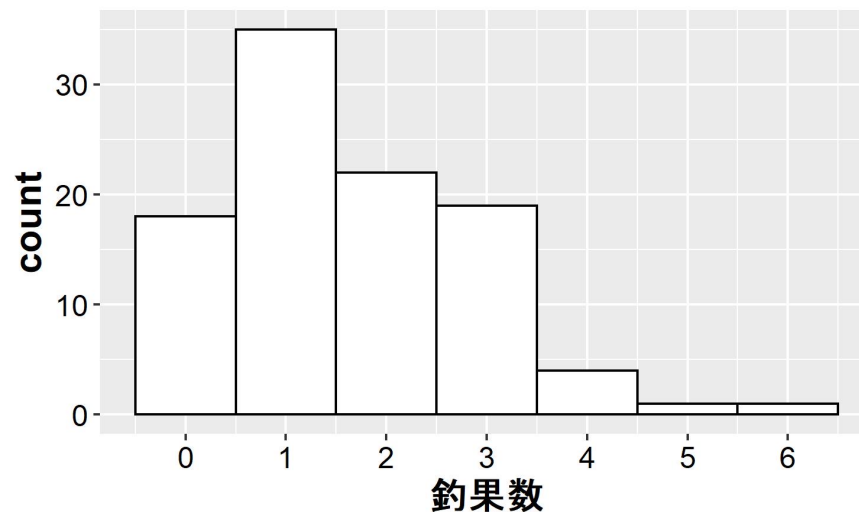
- ポアソン分布, 二項分布にしたがうデータを分析する
- リンク関数
- 尤度比検定と結果の解釈

正規線形モデルの限界

●正規分布を想定できないデータ

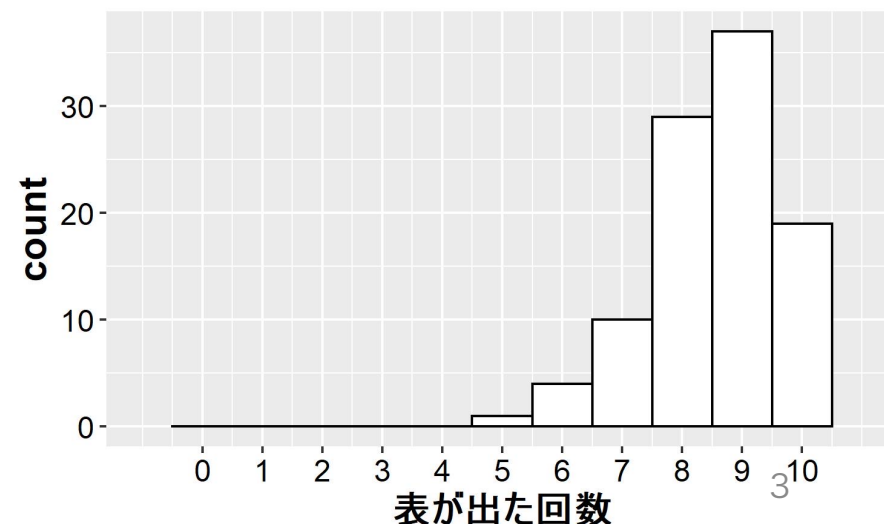
✓上限のないカウントデータ

- 例：釣りで釣れた魚の数
- 実測値は非負の整数
- 分布が左右非対称



✓上限のあるカウントデータ

- 例：コイントス10回中で表が出る回数
- 実測値は上限のある非負の整数
- 上限・下限に近いとき、分布が左右非対称



カウントデータなどでは、
正規分布を仮定することは不適切

正規分布以外も扱える線形モデルへ

●一般化線形モデル(generalized linear model : GLM)

- ✓正規分布以外の確率分布（ポアソン分布・二項分布・ガンマ分布など）も利用できる
- ✓データの性質（非負の実数, 0-1）にあわせて線形予測子を調整
- ✓正規線形モデルは, GLMで正規分布を想定した特殊ケース

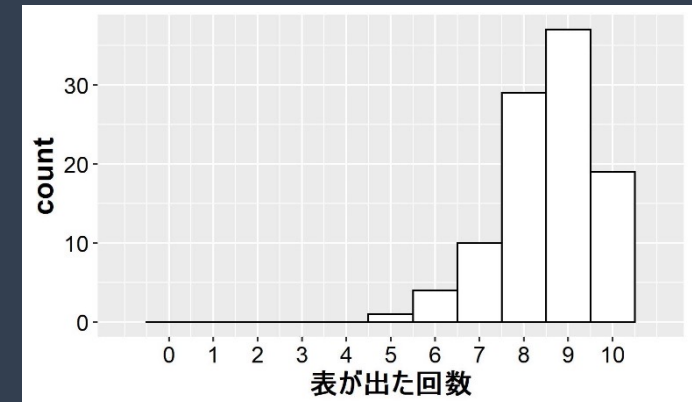
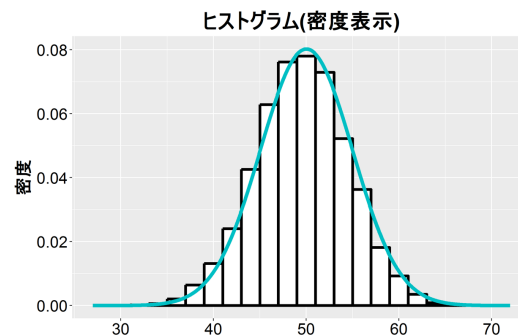
一般化線形モデル (GLM)

正規線形モデル

$$Y \sim \text{Normal}(\mu, \sigma)$$

$$\mu = \beta_0 + \beta_1 * d + \beta_2 * x + \dots$$

- 確率分布に正規分布を仮定
- 予測値の変動を直線で表現
- 応答変数は $-\infty$ から ∞ の範囲をとりうる連続変数



- 正規分布以外の確率分布も想定可能
 - ✓ ポアソン分布・二項分布etc.
- 応答変数にカウントデータも扱える

ポアソン分布の一般化線形モデル ポアソン回帰

ポアソン分布

●ポアソン分布 (Poisson distribution)

✓ある事象が生じた回数を表す確率分布

- 上限のないカウントデータが従う分布として有名

✓観測値は非負の整数

- 1時間あたりの流れ星の回数
- ある地域の動物の数
- 1 km^2 あたりにある木の本数

- 上限はないが、小さい値が比較的出やすい



ポアソン分布

●ポアソン分布 (Poisson distribution)

✓分布の平均と分散が等しい

- 平均=分散： λ (ラムダ)

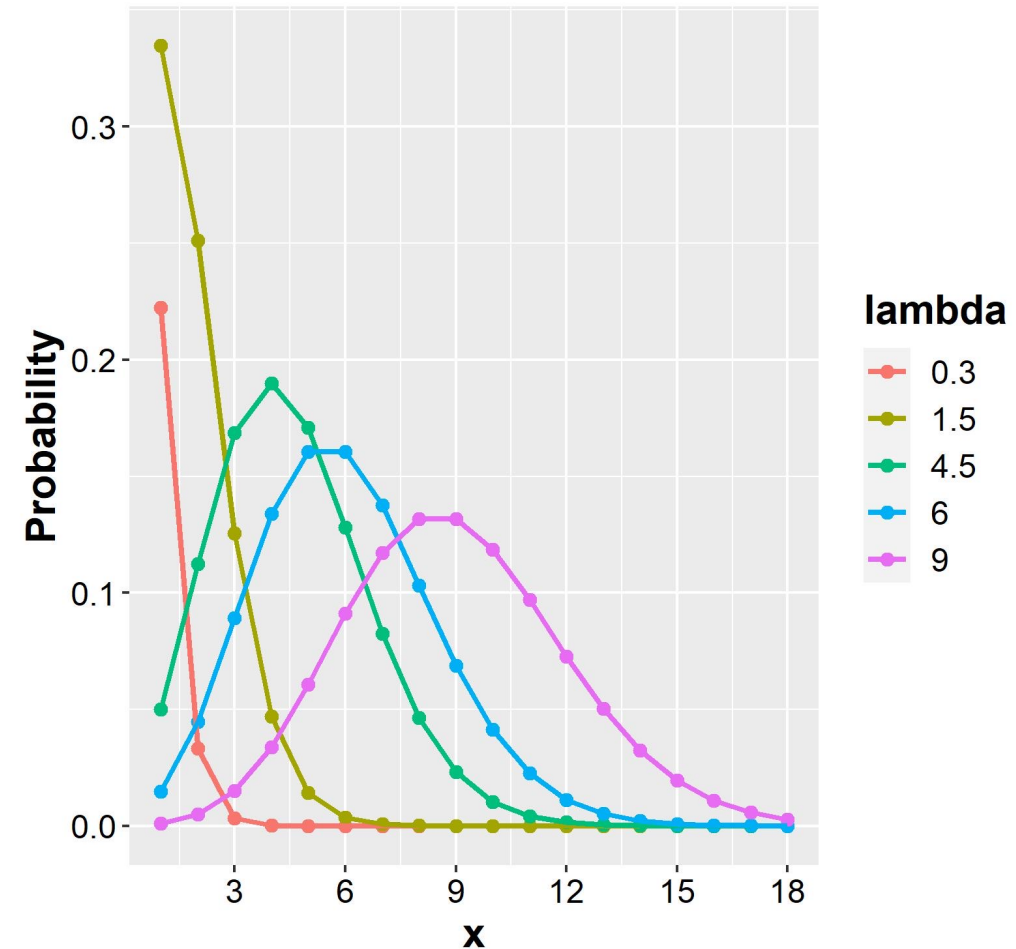
✓パラメータ λ の値で形状が変わる

- $Y \sim \text{Poisson}(\lambda)$

✓ Y ：応答変数の予測値 (非負の整数)

✓ λ ：平均および分散 (非負の実数)

✓平均が大きくなると形状が正規分布に近づく



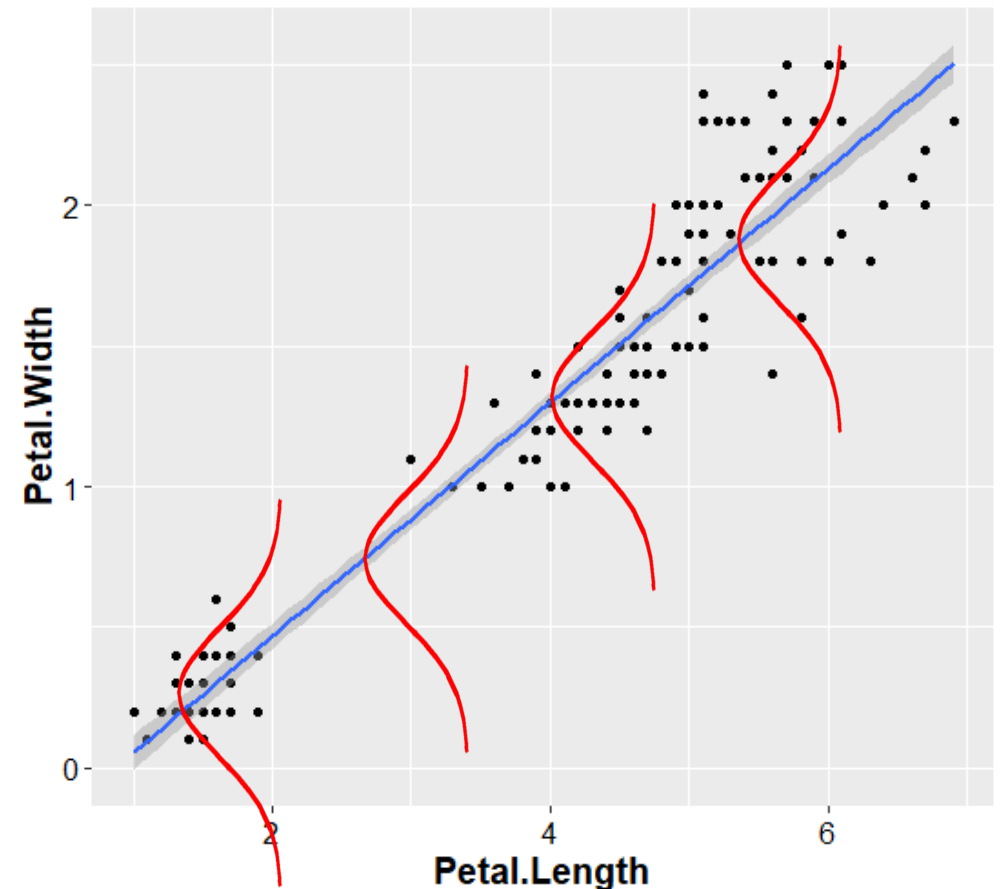
正規線形モデルの限界②

●予測値が直線で表現される

✓外挿していくと，予測値が $-\infty$ から ∞ の範囲までの値をとる

- 正規分布の平均も $-\infty \sim \infty$ に変化
- Petal.Widthの予測値は，
外挿していくと負の値をとりうる

●パラメータ λ ，観測値が非負の実数の値であるポアソン分布に対応できない



一般化線形モデルにおけるパラメータ推定

●ポアソン回帰モデル

- ✓確率分布にポアソン分布を仮定した統計モデル
 - ・説明変数によってポアソン分布のパラメータ λ が変動
 - ・logリンク関数で λ (正の実数)と線形予測子 $(-\infty \sim \infty)$ をつなぐ
- ✓線形予測子 z で多様なモデル構造を表現
 - ・質的変数も量的変数も
 - ・複数の説明変数・交互作用

$$Y \sim \text{Poisson}(\lambda)$$

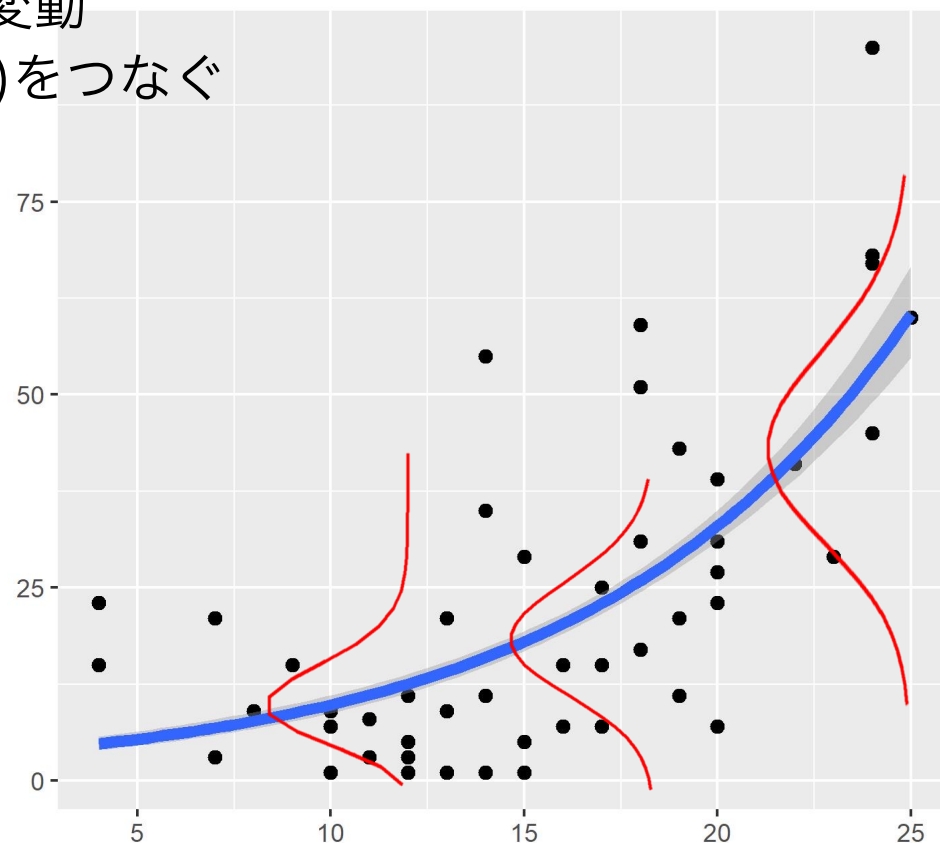
確率分布

$$\log \lambda = z$$

リンク関数

$$z = \beta_0 + \beta_1 x + \beta_2 d + \dots$$

線形予測子



リンク関数の導入

●リンク関数

- ✓線形予測子の値を確率分布のパラメータに変換する関数
- ✓ポアソン分布のパラメータの制約を満足するようなリンク関数

●logリンク関数

- ✓線形予測子 z ($-\infty \sim \infty$)とポアソン分布のパラメータ λ (正の実数)をつなぐ
- ✓対数関数なので、定義域は正の実数

$$Y \sim \text{Poisson}(\lambda)$$

確率分布

$$\log \lambda = z$$

リンク関数

$$z = \beta_0 + \beta_1 x + \beta_2 d + \dots$$

線形予測子

リンク関数の導入

●ログリンク関数の逆関数で線形予測子 z を λ に変換

✓対数関数の逆関数：指数関数

- 線形予測子 z が ∞ に近づくとき、 λ も ∞ に近づく
✓ポアソン分布の平均および分散が大きくなる
- 線形予測子 z が $-\infty$ に近づくとき、 λ は 0 に近づく
✓ポアソン分布の平均および分散が小さくなる

$$Y \sim \text{Poisson}(\lambda)$$

確率分布 ($\lambda > 0$)

$$\log \lambda = z$$

リンク関数



$$\lambda = \exp(z)$$

リンク関数の逆関数

$$z = \beta_0 + \beta_1 x + \beta_2 d + \dots$$

線形予測子 ($-\infty < z < \infty$)

一般化線形モデルでのパラメータ推定

「あてはまりの良さ」が最大になるようにパラメータが推定される

- 正規線形モデルにおけるあてはまりの指標

- ✓残差平方和

- 残差平方和が小さくなるパラメータを推定（最小二乗法）

- 一般化線形モデルにおけるあてはまりの指標

- ✓尤度 (Likelihood)

- あるパラメータを持つ確率分布において、観測値が得られる確率
- 尤度が大きいほどデータへのあてはまりが良い
- 尤度が最も大きくなるようなパラメータを推定（最尤法）

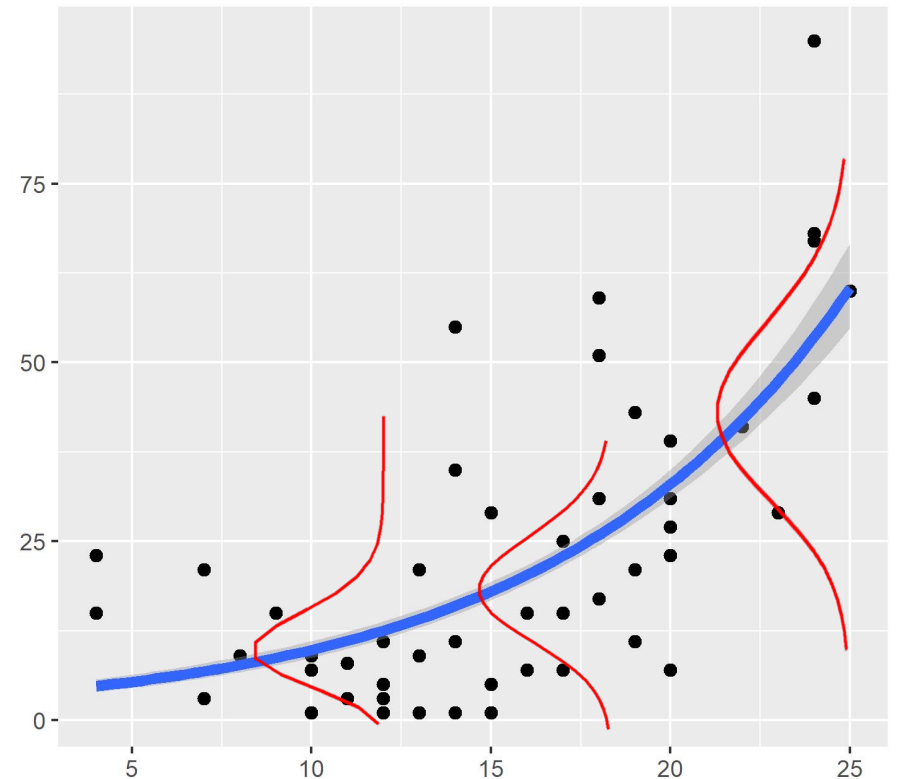
一般化線形モデルでのパラメータ推定

●最尤法

✓観測値から、尤度が最大になる確率分布のパラメータを推定

- 今回のデータが得られる確率が最大になるような確率分布のパラメータを求める
- ポアソン分布のパラメータ：平均=分散 λ
- 二項分布のパラメータ：事象の生起確率 q

●説明変数に伴う予測値（確率分布のパラメータ）の変動は線形予測子で表現



動物種の数ポアソン回帰でモデリング

●farawayパッケージのサンプルデータ"gala"

- ✓島の面積と、その島での動物種数のデータ
- ✓応答変数Speciesはカウントデータ（非負の整数）
 - 応答変数：Species（種数）
 - 説明変数：Area（島の面積）

```
gala2 <- gala %>%  
  mutate(area_log = log10(Area*100000))
```

```
library(faraway)  
head(gala)  
...
```

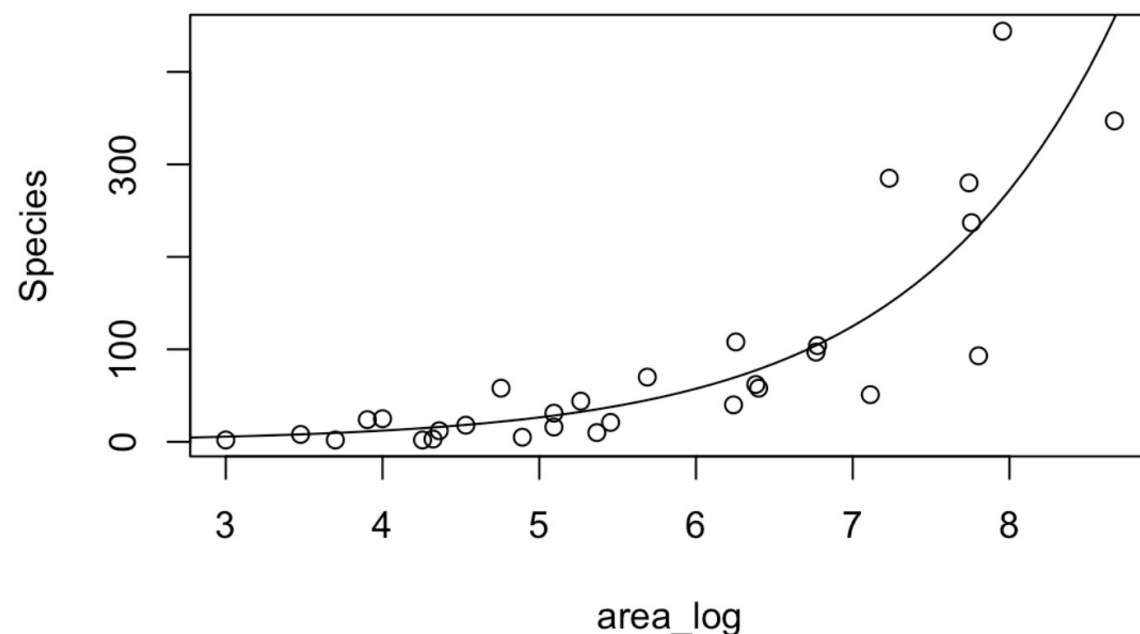
	Species <dbl>	Endemics <dbl>	Area <dbl>	Elevation <dbl>	Nearest <dbl>	Scruz <dbl>	Adjacent <dbl>
Baltra	58	23	25.09	346	0.6	0.6	1.84
Bartolome	31	21	1.24	109	0.6	26.3	572.33
Caldwell	3	3	0.21	114	2.8	58.7	0.78
Champion	25	9	0.10	46	1.9	47.4	0.18
Coamano	2	1	0.05	77	1.9	1.9	903.82
Daphne.Major	18	11	0.34	119	8.0	8.0	1.84

6 rows

動物種の数をもアソン回帰でモデリング

- Areaをそのまま使うと不具合があるので、単位をkm²からm²にして対数変換したarea_logを説明変数にした
 - ✓データフレーム名：gala2
 - ✓応答変数：Species（種数）
 - ✓説明変数：area_log（島の面積, m²）

```
gala2 <- gala %>%  
  mutate(area_log = log10(Area*100000))
```



動物種の数ポアソン回帰でモデリング

- 一般化線形モデルはglm()関数で

```
m_poisson <- glm(Species ~ area_log,  
                  family = poisson(link = "log"),  
                  data = gala2)
```

- ✓応答変数~説明変数の記法はlm()と同じ
- ✓引数「**family =**」で確率分布とリンク関数を指定
 - **family = poisson(link = "log")**でポアソン回帰
 - ポアソン回帰の場合はデフォルトでログリンク関数になる

- lm()の正規線形モデルと同様に, summary()でパラメータ推定
- Anova()で検定も可能

動物種のをポアソン回帰でモデリング

```
summary(m_poisson)
```

```
...
```

Call:

```
glm(formula = Species ~ area_log, family = poisson(link = "log"),  
     data = gala2)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-10.4688	-3.6073	-0.8874	2.9028	10.1517

Coefficients: 切片

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.61514	0.12066	-5.098	3.43e-07 ***
area_log	0.77767	0.01647	47.212	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 3510.73 on 29 degrees of freedom
Residual deviance: 651.67 on 28 degrees of freedom
AIC: 816.5

●パラメータの推定結果

$$Y \sim \text{Poisson}(\lambda)$$

$$\log \lambda = z$$

$$z = \beta_0 + \beta_1 d$$



$$z = -0.62 + 0.78x$$

predict()関数：予測値の計算関数

●predict()関数にモデルを入力すると予測値を計算できる

✓引数：fittedオブジェクト

✓返り値：線形予測子のベクトル

- データフレームの各行での説明変数の値における予測値を計算
- 返り値の数値の順序は、元のデータフレームの行の順序と一致

```
predict(m_poisson)
```

$$Y \sim \text{Poisson}(\lambda)$$

$$\log \lambda = z$$

$$z = \beta_0 + \beta_1 d + \beta_2 x + \beta_3 dx$$

Baltra	Bartolome	Caldwell	Champion	Coamano	Daphne.Major
4.361548	3.345851	2.746111	2.495531	2.261429	2.908846
Daphne.Minor	Darwin	Eden	Enderby	Espanola	Fernandina
2.420167	3.558881	2.088904	2.694048	4.646131	5.452558
Gardner1	Gardner2	Genovesa	Isabela	Marchena	Onslow
3.083351	3.189285	4.236964	6.126664	4.915820	1.717862
Pinta	Pinzon	Las.Plazas	Rabida	SanCristobal	SanSalvador
4.653527	4.248447	2.776835	3.809254	5.405288	5.417736
SantaCruz	SantaFe	SantaMaria	Seymour	Tortuga	Wolf
5.572053	4.347671	5.009573	3.479140	3.345851	3.626918

predict()関数：予測値の計算関数

●type = “response”オプションでGLMにも対応

✓predict(…, type = “response”)

- 引数：fittedオブジェクト
- 返り値：確率分布のパラメータのベクトル

$Y \sim \text{Poisson}(\lambda)$

$\log \lambda = z$

$z = \beta_0 + \beta_1 d + \beta_2 x + \beta_3 dx$

```
predict(m_poisson, type = "response")
```

Baltra	Bartolome	Caldwell	Champion	Coamano	Daphne.Major
78.378356	28.384718	15.581913	12.128171	9.596796	18.335618
Daphne.Minor	Darwin	Eden	Enderby	Espanola	Fernandina
11.247738	35.123872	8.076062	14.791437	104.181164	233.354367
Gardner1	Gardner2	Genovesa	Isabela	Marchena	Onslow
21.831443	24.271070	69.197474	457.906022	136.431146	5.572602
Pinta	Pinzon	Las.Plazas	Rabida	SanCristobal	SanSalvador
104.954478	69.996596	16.068091	45.116760	222.580261	225.368202
SantaCruz	SantaFe	SantaMaria	Seymour	Tortuga	Wolf
262.973318	77.298222	149.840772	32.431830	28.384718	37.596781

predict()関数：予測値の計算関数

- 線形予測子 z をリンク関数の逆関数に投入すると，パラメータ λ

```
> z <- predict(m_poisson)[1:5]
> exp(z)[1:5]
  Baltra Bartolome Caldwell Champion Coamano
78.378356 28.384718 15.581913 12.128171  9.596796
>
>
> predict(m_poisson, type = "response")[1:5]
  Baltra Bartolome Caldwell Champion Coamano
78.378356 28.384718 15.581913 12.128171  9.596796
```

$Y \sim \text{Poisson}(\lambda)$ 確率分布 ($\lambda > 0$)

$\log \lambda = z$ リンク関数 $\longleftrightarrow$ $\lambda = \exp(z)$ リンク関数の逆関数

$z = \beta_0 + \beta_1 x + \beta_2 d + \dots$ 線形予測子 ($-\infty < z < \infty$)

モデルの回帰係数の解釈

- 切片：説明変数がすべて0をとる場合の線形予測子

$$Y \sim \text{Poisson}(\lambda)$$

$$\log \lambda = z$$

$$z = -0.62 + 0.78x$$

- ✓説明変数の値が0のときの線形予測子は

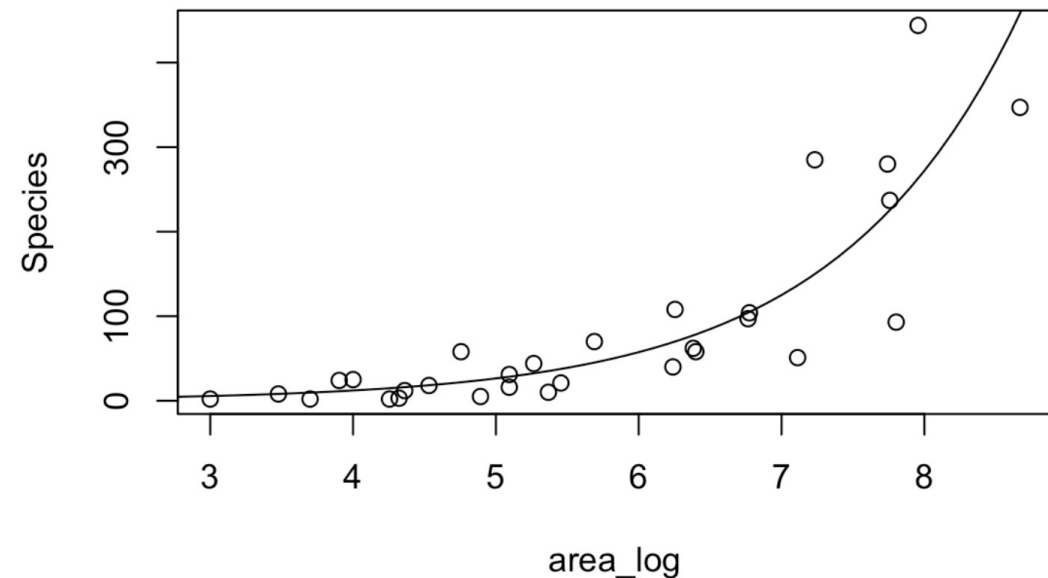
$$z = \beta_0 = -0.62$$

- ✓ポアソン分布のパラメータ λ は

$$\lambda = \exp(\beta_0)$$

- ✓ $\exp(-0.62) = 0.54$

- ✓面積が0のときの種数の予測は平均 0.54



モデルの回帰係数の解釈

- 傾き：説明変数が1大きくなったときの線形予測子の変化量

$$Y \sim \text{Poisson}(\lambda)$$

$$\log \lambda = z$$

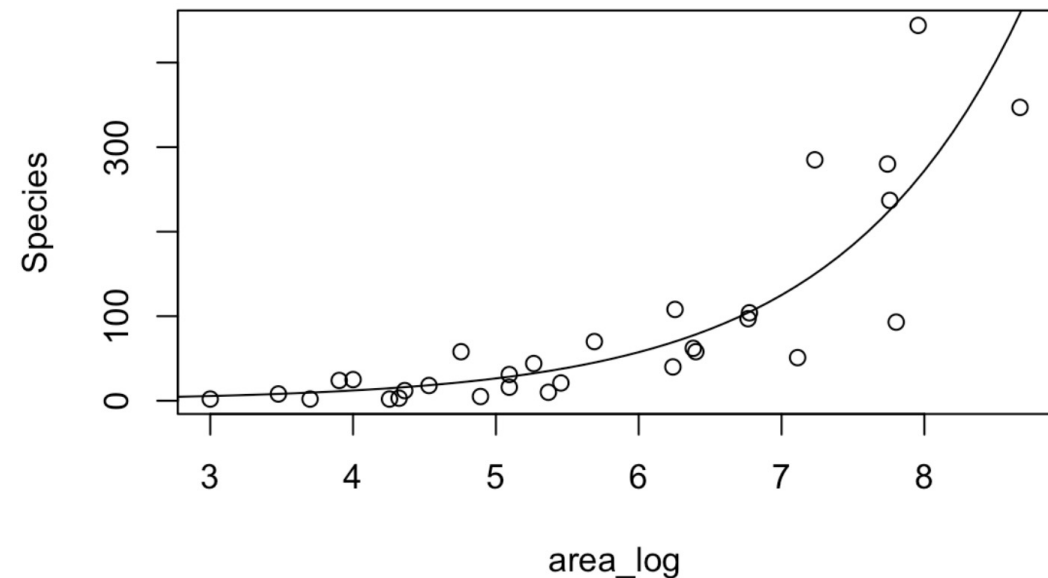
$$z = -0.62 + 0.78x$$

✓説明変数の値が1大きくなったときの線形予測子 z の変化量：0.78

✓ x が1増えると， λ は $\exp(0.78)$ 倍になる

$$\exp(0.78) = 2.18$$

✓面積が1m²増えると，種数は2.18倍



Anova()で尤度比検定

- `car::Anova()`関数に，説明変数の最も多いモデルを投入
 - ✓ 尤度比検定の検定統計量はカイ二乗値

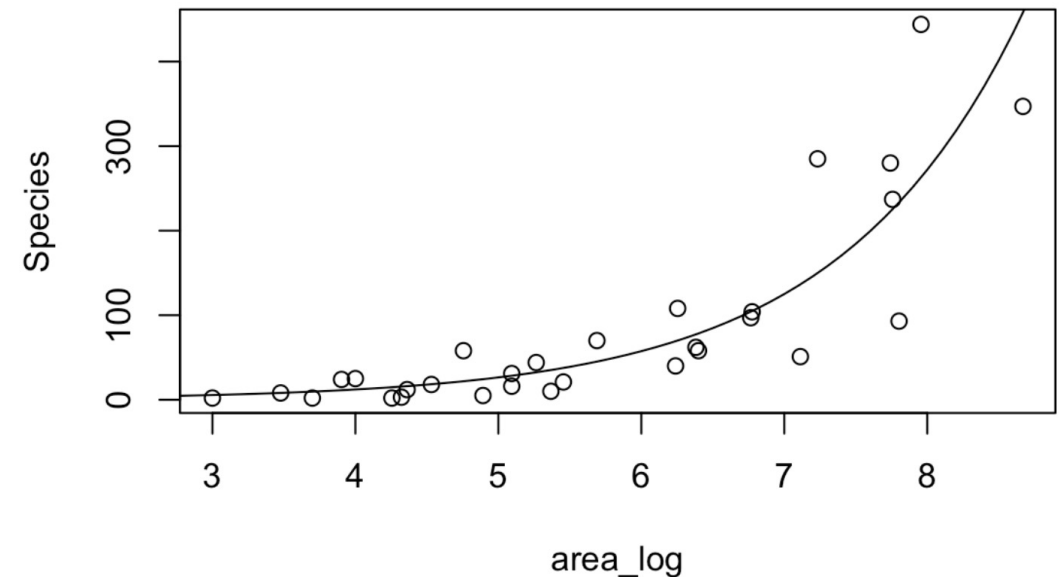
```
Anova(m_poisson)
```

```
Analysis of Deviance Table (Type II tests)
```

```
Response: Species
```

	LR	Chisq	Df	Pr(>Chisq)
area_log	2859.1	1	< 2.2e-16	***

```
---
```



尤度比検定の結果，面積の主効果が有意であった ($\chi^2(1) = 2859.1, p < .001$)。島の面積が大きいほど動物の種数が多かった。

●対数尤度 (log likelihood)

- ✓尤度の対数をとった値。元の尤度の大小関係は保たれる
 - モデルのあてはまりの良さを比較するとき、対数尤度が用いられる
 - logLik()関数にfittedオブジェクトを投入すると対数尤度が計算できる
- ✓Anova()出力のLR列の値は、2つのモデルの対数尤度の差分×2
 - 説明変数area_logを入れたモデルは、入れないモデルより、対数尤度がこれだけ増えた

```
Anova(m_poisson)
...
```

Analysis of Deviance Table (Type II tests)

Response: Species

	LR Chisq	Df	Pr(>Chisq)
area_log	2859.1	1	< 2.2e-16 ***

二項分布の一般化線形モデル ロジスティック回帰

二項分布

●二項分布 (Binomial distribution)

✓二値が得られる独立試行 N 回で、一方の値が生じた回数を表す確率分布

- 上限のあるカウントデータの分布
- 観測値は非負の整数（上限は N ）

✓各試行が独立であることが前提

- 前試行の結果が次の試行に影響しない（コイントス）
- 二値データ（0/1, はい/いいえ）

✓ある事象が起こる回数の割合を扱う場合に用いることが多い

- 10回のコイントスで表が出た回数（実際にやってみよう）
- 30人のクラスにおける女性の人数



二項分布

●二項分布 (Binomial distribution)

✓事象が生じる確率 q , 試行回数の上限 N

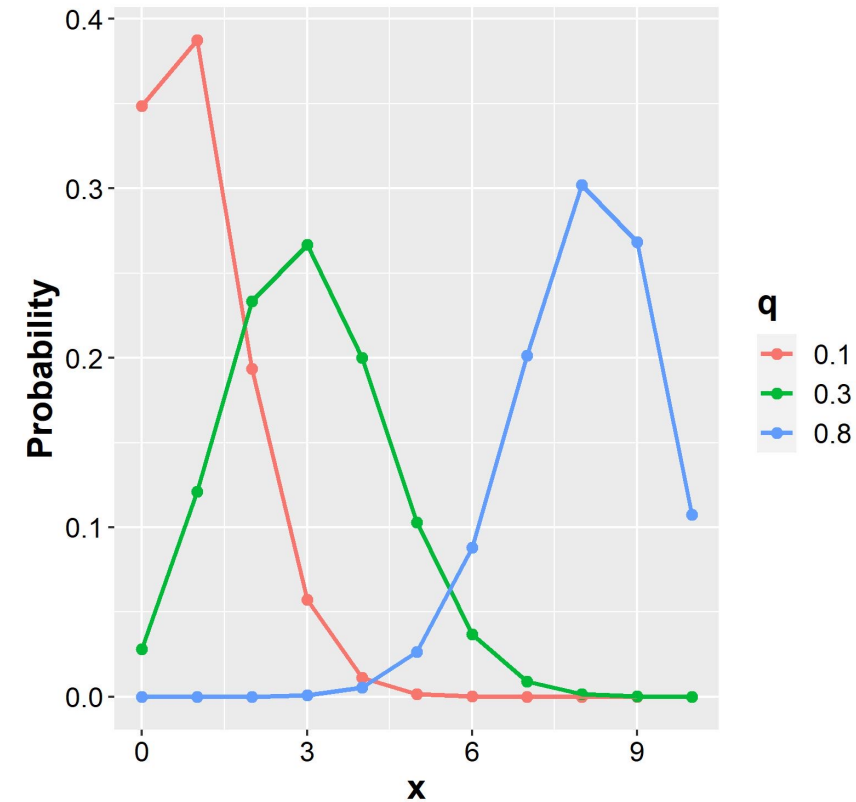
• $Y \sim \text{Binomial}(N, q)$

✓ q : 一方の事象が生じる確率 (0~1の実数)

✓ $1 - q$: もう一方の事象が生じる確率(0~1の実数)

✓ N : 試行回数の上限 (非負の整数)

✓ 上限・下限に近付くほど, 分布の形状が歪む



二項分布を仮定した一般化線形モデル

●ロジスティック回帰モデル

✓確率分布に二項分布を仮定した統計モデル

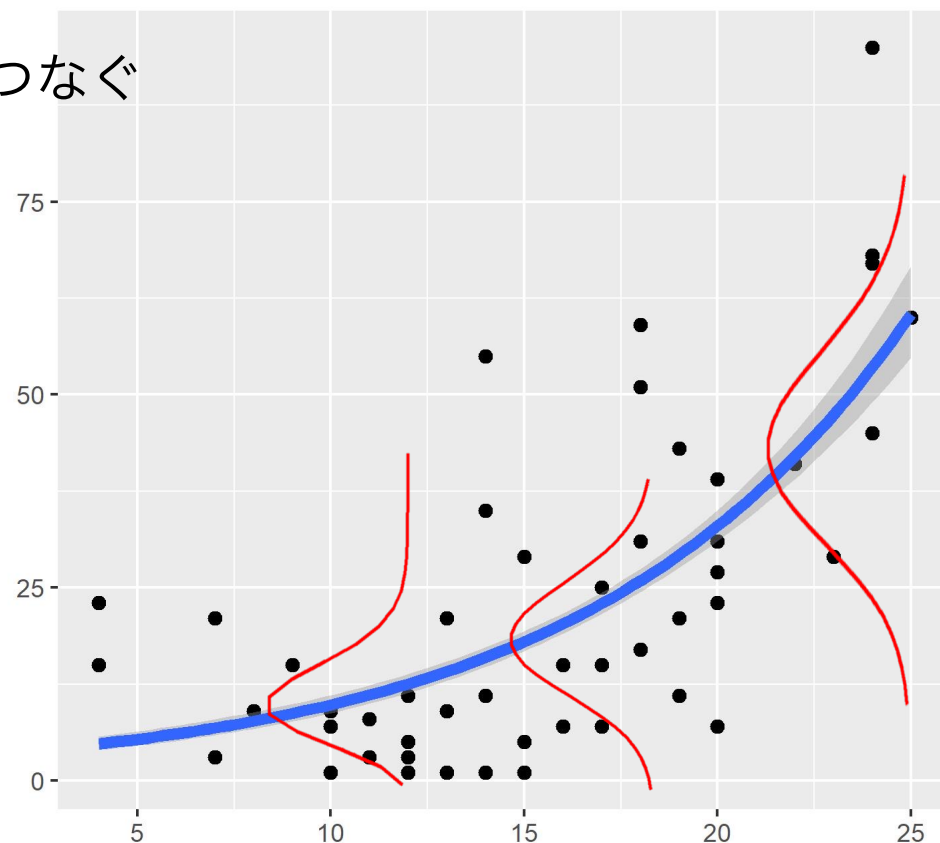
- 説明変数によって二項分布のパラメータ q が変動
- logitリンク関数で q (0-1)と線形予測子 $(-\infty \sim \infty)$ をつなぐ

✓線形予測子 z で多様なモデルを表現

$Y \sim \text{Binomial}(N, q)$ 確率分布

$\text{logit } q = z$ リンク関数

$z = \beta_0 + \beta_1 x + \beta_2 d + \dots$ 線形予測子



リンク関数の導入

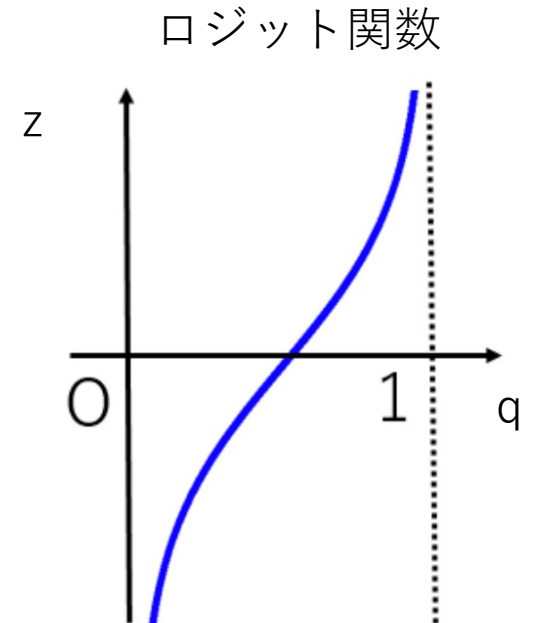
●二項分布の場合は、ロジットリンク関数

- ✓線形予測子 z ($-\infty \sim \infty$)と二項分布のパラメータ q ($0-1$)をつなぐ
- ✓ロジット関数の定義域は0から1の実数

$Y \sim \text{Binomial}(N, q)$ 確率分布

$\text{logit } q = z$ リンク関数 (ロジット関数)

$z = \beta_0 + \beta_1 x + \beta_2 d + \dots$ 線形予測子

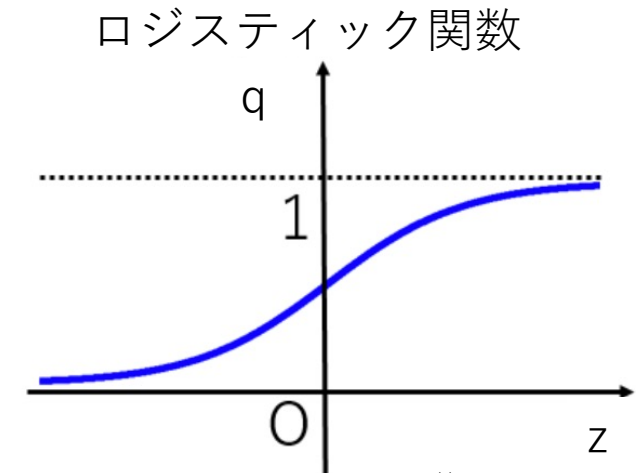


リンク関数の導入

●リンク関数（ロジット）の逆関数で線形予測子 z を q に変換

✓ロジット関数の逆関数：ロジスティック関数（値域が0-1）

- z が ∞ に近づくとき, q は 1 に近づく
 - ✓ 事象の生起確率が大きくなる
- z が $-\infty$ に近づくとき, q は 0 に近づく
 - ✓ 事象の生起確率が小さくなる



<https://mathwords.net/logitkansu>

$$Y \sim \text{Binomial}(N, q)$$

確率分布 ($0 < q < 1$)

ロジスティック関数

$$\text{logit } q = z$$

ロジット関数

$$q = \text{logistic}(z) = \frac{1}{1 + \exp(-z)}$$

$$z = \beta_0 + \beta_1 x + \beta_2 d + \dots$$

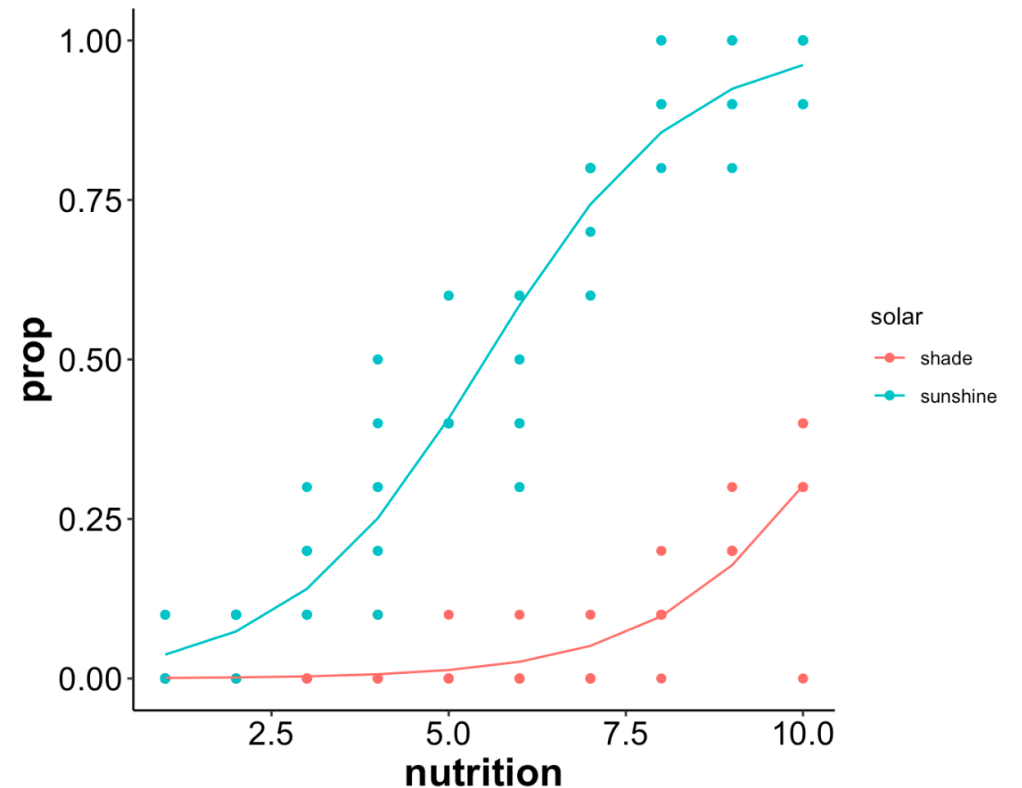
線形予測子 ($-\infty < z < \infty$)

種子の発芽確率をロジスティック回帰でモデリング

- 10個の種子が発芽した数のデータ seed.csv
 - ✓ germination : 発芽数 (上限のあるカウントデータ)
 - ✓ size : 種子数 = 10
 - ✓ solar : 日照条件 (shade, sunshine)
 - ✓ nutrition : 栄養素の量 (1~10の整数)

```
seed <- read.csv("seed.csv")  
head(seed)
```

	germination	size	solar	nutrition	prop
1	0	10	shade	1	0
2	0	10	shade	1	0
3	0	10	shade	1	0
4	0	10	shade	1	0
5	0	10	shade	1	0
6	0	10	shade	2	0



種子の発芽確率をロジスティック回帰でモデリング

- 応答変数：10個中の発芽数：`cbind(分子, 分母-分子)`
- 説明変数：`solar`と`nutrition`の主効果および交互作用：`solar*nutrition`
- 一般化線形モデル`glm()`で記述

```
m_binom <- glm(cbind(germination, size-germination) ~ solar*nutrition,  
               family = binomial(link = "logit"),  
               data = seed)
```

- ✓ 引数「**family =**」で確率分布とリンク関数を指定
 - `family = binomial(link = "logit")`でロジスティック回帰
 - 二項分布, ロジットリンク関数

パラメータの推定結果を確認

```
summary(m_binom)
```

Call:

```
glm(formula = cbind(germination, size - germination) ~ solar *  
    nutrition, family = binomial(link = "logit"), data = seed)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.6823	-0.5226	-0.1556	0.4187	1.9098

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-7.78852	1.09633	-7.104	1.21e-12	***
<u>solarsunshine</u>	<u>3.82173</u>	1.15083	3.321	0.000897	***
<u>nutrition</u>	<u>0.69514</u>	0.12434	5.590	2.26e-08	***
<u>solarsunshine:nutrition</u>	<u>0.02333</u>	0.13779	0.169	0.865526	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 656.621 on 99 degrees of freedom
Residual deviance: 68.545 on 96 degrees of freedom
AIC: 202.9

●パラメータの推定結果

$$Y \sim \text{Binomial}(N, q)$$

$$\text{logit } q = z$$

$$z = \beta_0 + \beta_1 d + \beta_2 x + \beta_3 dx$$



$$z = -7.79 + 3.82d + 0.69x + 0.02dx$$

モデルの回帰係数の解釈

●切片：説明変数がすべて0をとる場合の線形予測子

$$Y \sim \text{Binomial}(N, q)$$

$$\text{logit } q = z$$

$$z = -7.79 + 3.82d + 0.69x + 0.02dx$$

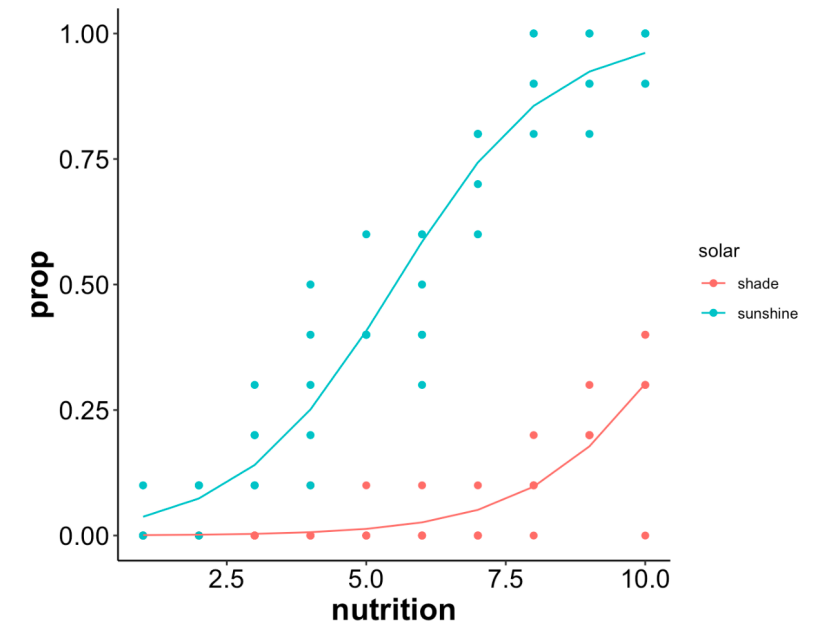
✓説明変数が全て0のときの線形予測子 z の値は β_0

$$z = \beta_0 = -7.79$$

✓二項分布のパラメータ q は、ロジスティック関数に従い $q = 1/(1 + \exp(-z))$

✓ $q = 1/(1 + \exp(7.79)) = 0.0004$

✓日照がshadeでnutritionが0のとき、発芽率は0.0004



モデルの回帰係数の解釈

●傾き：説明変数が1大きくなったときの線形予測子の変化

$$Y \sim \text{Binomial}(N, q)$$

$$\text{logit } q = z$$

$$z = -7.79 + 3.82d + 0.69x + 0.02dx$$

- ✓日照条件は固定で、栄養素が1大きくなったとき、線形予測子 z の変化量は β_2
 - ✓solar = shadeで、nutritionが1増えると、オッズ $\frac{q}{1-q}$ が $\exp(\beta_2)$ 倍になる
 - ✓ $\exp(0.69) = 1.99$
 - ✓日陰条件では、栄養素が1増えるとオッズは 1.99倍

●オッズ $\frac{q}{1-q}$ ：ある事象が起こる確率と起こらない確率の比

- ✓事象が起こる確率が高い場合、オッズは ∞ に近づく
- ✓事象が起こる確率が低い場合、オッズは0に近づく
- ✓起こる確率が0.5のときは、オッズは1

- sloar=shadeにおいて（ダミー変数d=0），オッズを線形予測子zで表すと，

$$\begin{aligned}\frac{q}{1-q} &= \left(\frac{1}{1 + \exp(-z)} \right) / \left(\frac{\exp(-z)}{1 + \exp(-z)} \right) \\ &= \exp(z) \\ &= \exp(\beta_0 + \beta_2 x)\end{aligned}$$

- 栄養素が0のとき， $x=0$ なので， $\frac{q}{1-q} = \exp(\beta_0)$

- 栄養素が1のとき， $\frac{q}{1-q} = \exp(\beta_0 + \beta_2)$
 $= \exp(\beta_0) * \exp(\beta_2)$

となり，オッズは栄養素が0のときの $\exp(\beta_2)$ 倍になるのがわかる

Anova()で尤度比検定

- car::Anova()関数に，説明変数の最も多いモデルを投入

```
Anova(m_binom)
```

```
Analysis of Deviance Table (Type II tests)
```

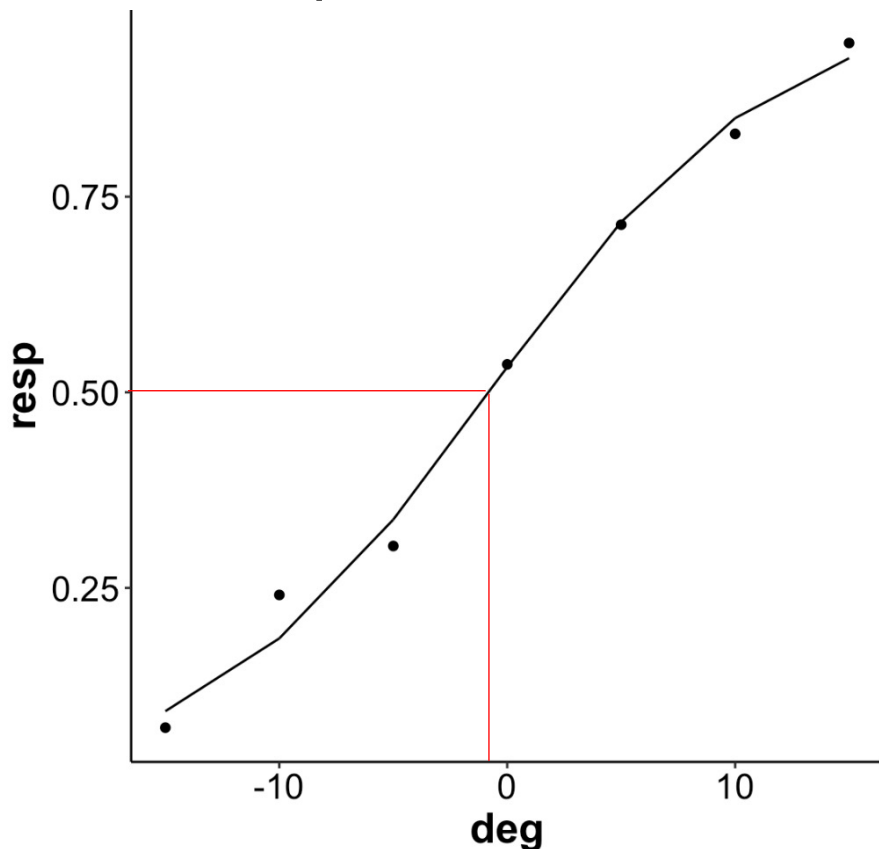
```
Response: cbind(germination, size - germination)
```

	LR	Chisq	Df	Pr(>Chisq)
solar	349.44	1		<2e-16 ***
nutrition	338.09	1		<2e-16 ***
solar:nutrition	0.03	1		0.8663

尤度比検定の結果， solarの主効果と ($\chi^2(1) = 349.44$, $p < .001$), nutritionの主効果が有意であったが($\chi^2(1) = 338.09$, $p < .001$)， 交互作用は有意でなかった($\chi^2(1) = 0.03$, $p = .866$)。発芽率は日照がある時のはない時よりも高く， 栄養素が増えると高くなった。

GLMを利用した閾値の推定

- 恒常法の実験データからPSE（主観的等価点）を推定したい
 - ✓ deg : $-15^{\circ} \sim 15^{\circ}$ の線分が右に傾いて見えたか, 左に傾いて見えたか
 - ✓ resp : 右=1, 左=0の二値データ



```
pse <- function(df,x,y){  
  c <- coef(glm(y ~ x,  
                family = binomial(link = "logit"),  
                data = df));  
  p <- c(-c[1]/c[2], 1/c[2])  
  return(p[1])  
}
```

```
> pse(d_pse, d_pse$deg, d_pse$resp)  
(Intercept)  
-0.803209
```

補足：パッケージでもできる

●quickpsyパッケージのquickpsy()関数

開発者Daniel Linares先生のHPに詳しい説明
<http://dlinares.org/quickpsy.html>

```
library(quickpsy)

fit <- quickpsy(d_pse, deg, resp,
               grouping = c("sub")
               )
```

quickpsy(df, 説明変数, 応答変数,
grouping = c(グルーピングしたい変数))

fit

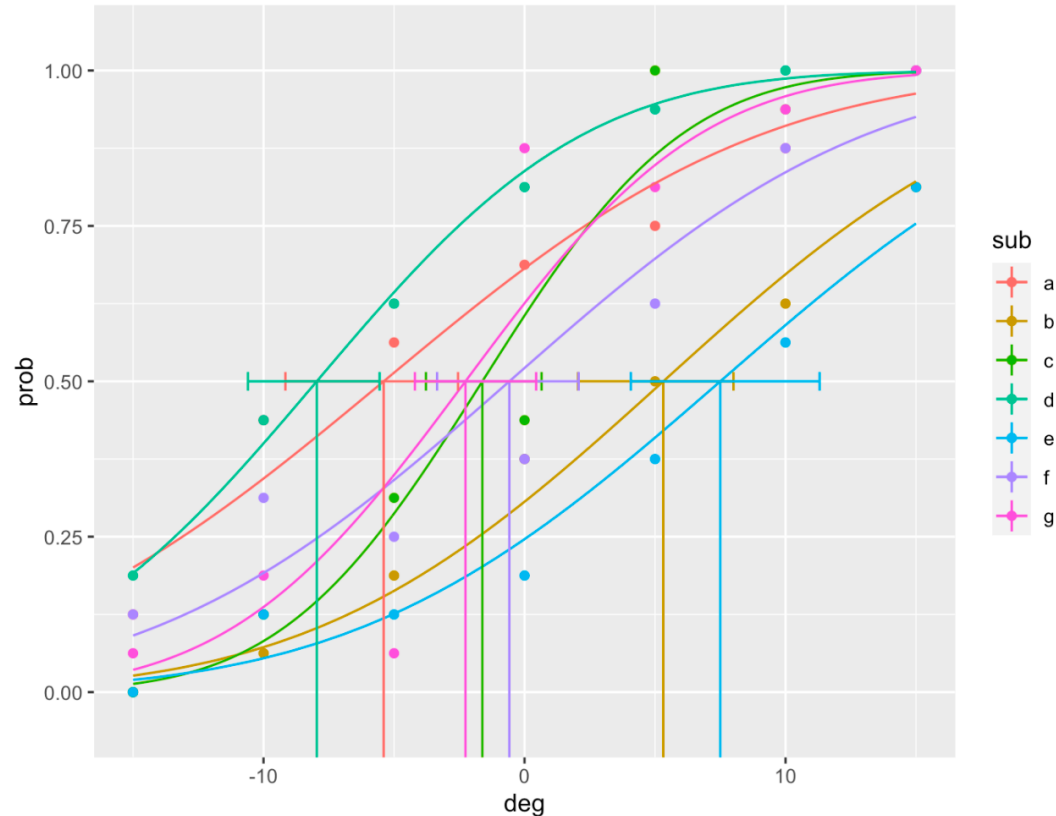
sub	thre	prob	threinf	thresup
a	-5.396865	0.5	-8.666642	-2.4019546
b	5.317894	0.5	3.066599	9.3254797
c	-1.616685	0.5	-3.274535	0.2003844
d	-7.959717	0.5	-10.646169	-5.7387053
e	7.501853	0.5	4.691691	11.1708665
f	-0.580823	0.5	-3.562570	1.6736680
g	-2.263665	0.5	-4.442833	0.0289520

- ✓ グループごとの閾値 (thre) などを出してくれる
 - グループ数によっては結構重いので注意

補足：パッケージでもできる

●quickpsyパッケージのquickpsy()関数

```
plot(fit, color=sub)
```



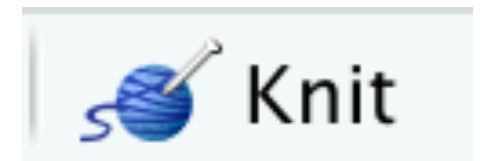
✓ quickpsy()関数の出力をplot()関数に入れるだけで、ggplotの図まで出してくれる

一般化線形モデルで各種分析

- 線形モデルの枠組みで、様々な分析が統一的に扱える
 - ✓線形予測子, リンク関数, 確率分布
 - ✓扱いたいデータが従う確率分布に応じたモデルを構築
 - ✓データフレームはロング型の整然 (tidy) データを用意すること
- さらに, ランダム効果を導入した一般化線形混合モデルへの拡張
 - ✓lme4パッケージのlmer()やglmer()関数で実装

今回紹介した主なパッケージおよび関数

- 正規線形モデル `lm()`関数
- 一般化線形モデル `glm()`関数
- モデルの要約 `summary()`関数
- F検定・尤度比検定 `car::Anova()`関数
- 多重比較 `multcomp::glht()`, `emmeans::emmeans()`



RMarkdownをknitして
htmlを出力しましょう

参考資料

- 統計モデリングおよび尤度比検定の詳しい説明

- ✓久保拓弥（2012）岩波書店

- 『データ解析のための統計モデリング入門

- ー一般化線形モデル・階層ベイズモデル・MCMCー』



- 一般化線形混合モデルの説明とRでの実装

- ✓堀裕亮（2017）ナカニシヤ書店

- 『ゼロからはじめる統計モデリング』

