GLMの応用範囲を広げる

XIAO LIYING

6.1 さまざまな種類のデータで応用できるGLM

表 6.1　R 内で GLM の構築に使える確率分布の一部．一般化線形モデル推定関数 glm() の family 指定と，よく使うリンク関数．対数リンク関数については第 3 章を，ロジットリンク関数についてはこの章の 6.4 節を参照．

	確率分布	乱数生成	glm() の family 指定	よく使うリンク関数
(離散)	二項分布	rbinom()	binomial	logit
	ポアソン分布	rpois()	poisson	log
	負の二項分布	rnbinom()	(glm.nb() 関数)	log
(連続)	ガンマ分布	rgamma()	gamma	log かな？
	正規分布	rnorm()	gaussian	identity

6.2 例題 : 上限のあるカウントデータ

架空植物の個体 i
観察種子数 Ni
発芽能力がある種子数 yi
死んだ種子数 Ni-yi
個体iの生存確率 qi
個体サイズ xi

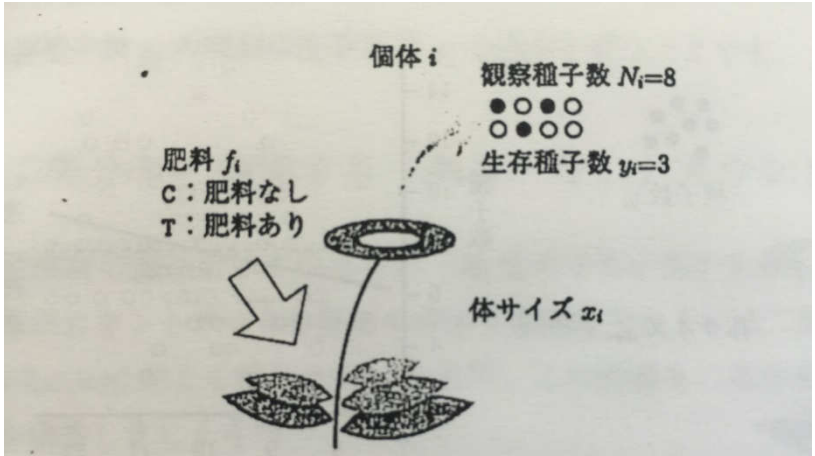
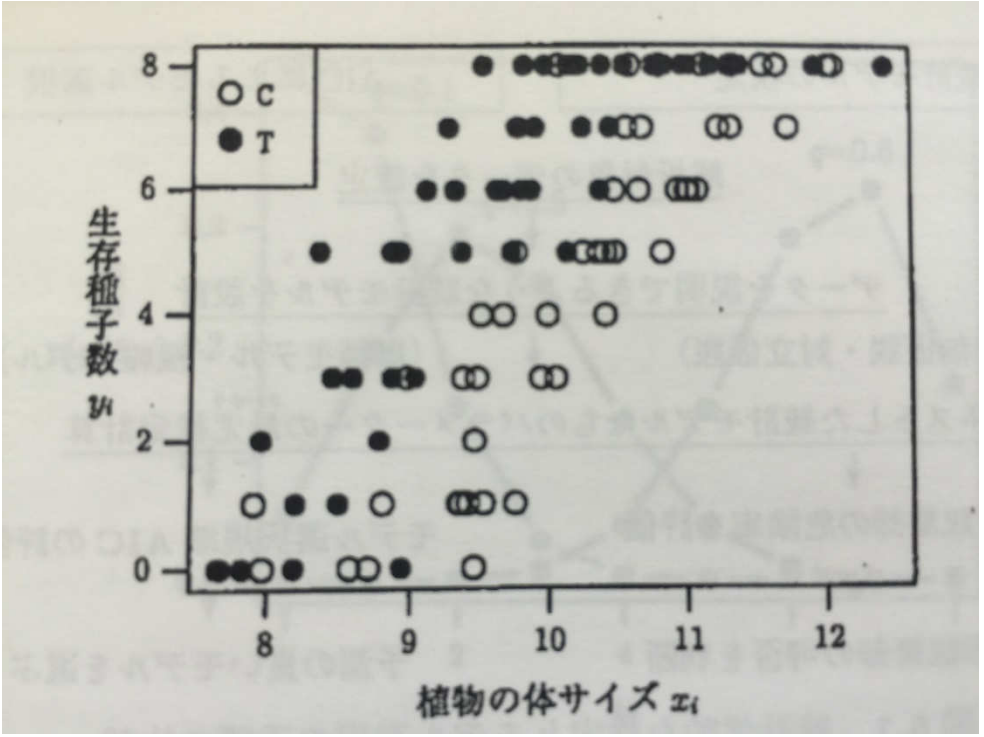


図 6.1　架空植物の第 i 番目の個体．この植物の体サイズ（個体の大きさ）x_i と施肥処理 f_i が種子の生存確率にどう影響しているのかを知りたい．

ある個体の生存確率qiがサイズxiや施肥処理fiといった説明変数によって、どう変化する

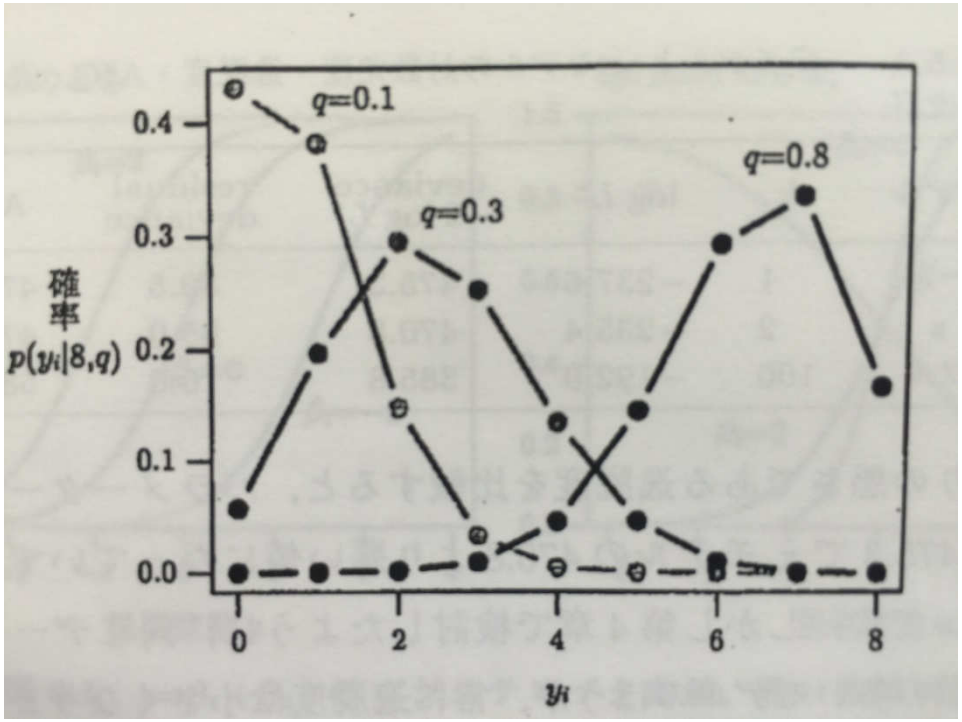
6.2 例題：上限のあるカウントデータ

これを見ると、
- 体サイズxiが大きくなると生存種子数yiが多くなるらしい
- 肥料をやると（fi=T）生存種子数yiが多くなるらしい



6.3 二項分布で表現する「あり、なし」カウントデータ

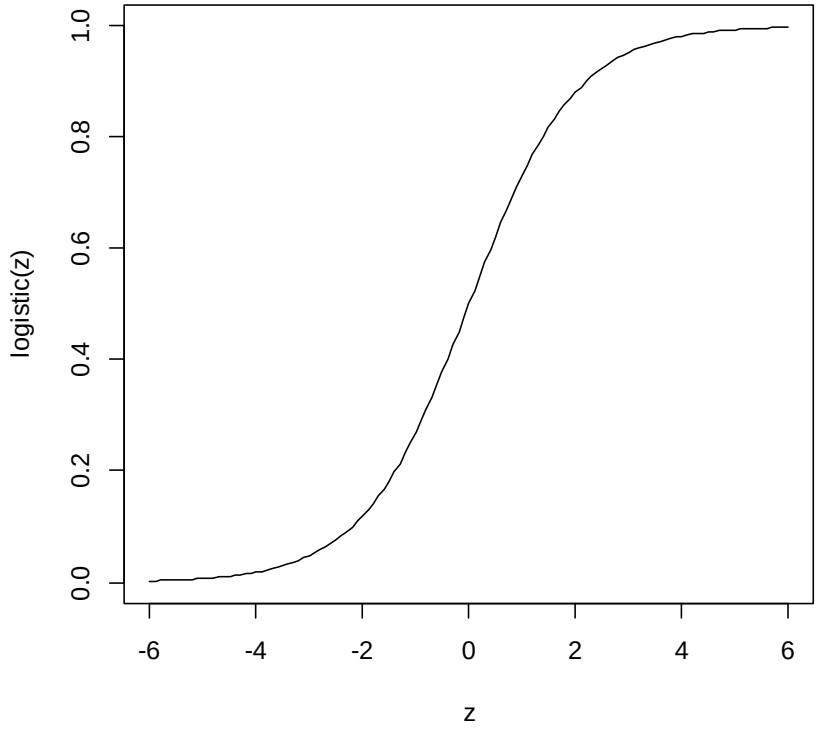
- 二項分布の確率分布は　$\mathrm{p}(y|N,q)=\binom{N}{y}q^{y}(1-q)^{N-y}$
- $\mathrm{p}(y|N,q)$ ==⇒N個中のy個で事項が生起する確率
- $\binom{N}{y}$➔場合の数



6.4 ロジスティック回帰とロジットリンク関数

- ロジスティック回帰では確率分布は二項分布、リンク関数はロジットリンク関数を指定します。
- ロジスティック関数の関数形は$q_i = logistic(z_i) = \frac{1}{1+\exp(-z_i)}$
- $z_i = \beta_1 + \beta_2 x_i + \cdots$

ロジスティック曲線➔



6.4 ロジスティック回帰とロジットリンク関数

- パラメーター推定
- 尤度関数 $L(\{\beta_j\}) = \prod_i \binom{N_i}{y_i} q^{y_i}(1-q_i)^{N_i - y_i}$
- >glm(cbind(y,N-y)~x+f,data=d,family=binomial)
- Coefficients:
- (Intercept)　　x　　fT
- -19.536　　1.952　　2.022

$\{\beta 1, \beta 2, \beta 3\}$={-19.5,1.95,2.02}

6.4 ロジスティック回帰とロジットリンク関数

- ロジットリンク関数の意味、解析
- $logit(q_i) = log\,\frac{qi}{1-qi}$
- ➔$\frac{qi}{1-qi} =$ exp(線形予測子)
- $=$exp($\beta 1 + \beta 2xi + \beta 3fi$)
- 左辺はオッズと呼ばれる量です

6.5 交互作用項の入った線形予測子

- $logit(q_i) = \beta_1 + \beta_2 x_i + \beta_3 f_i + \beta_4 x_i f_i$
- glm(cbind(y,N-y)~x*f,family=binomial,data=d)

- Call: glm(formula = cbind(y, N - y) ~ x * f, family = binomial, data = d)

- Coefficients:
- (Intercept) x fT x:fT
- -18.52332 1.85251 -0.06376 0.21634

- Degrees of Freedom: 99 Total (i.e. Null); 96 Residual
- Null Deviance: 499.2
- Residual Deviance: 122.4 AIC: 273.6

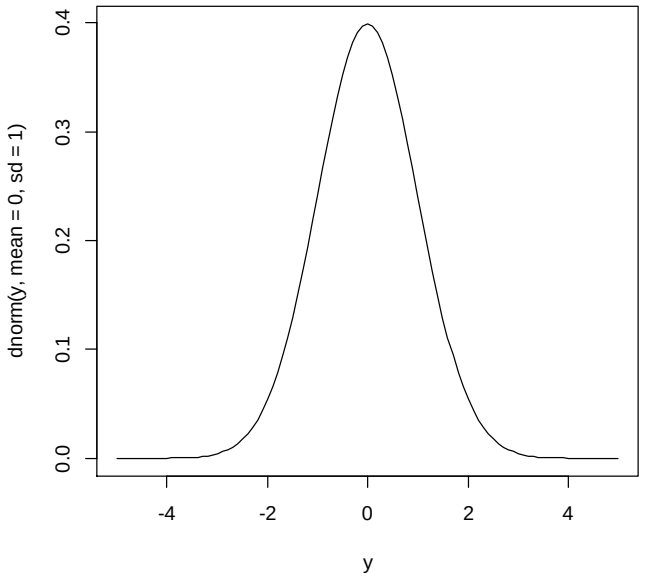
6.6 割算値の統計モデリングはためよう

- $\frac{\text{平均個体数}\lambda i}{Ai} =$ 人口密度
- $\lambda i = Ai *$ 人口密度 $= Aiexp(\beta 1 + \beta 2xi)$
- =exp($\beta 1 + \beta 2xi +$ $\boldsymbol{logAi}$)
- > glm(y~x,offset=log(A),family=poisson,data=d)

- Coefficients:
- (Intercept)　　x
- 　0.9731　　1.0383
- 個体数を調査地面積で割って密度にする、と言った観測値どうしの割算はまったく不要です。

6.7 正規分布とその尤度

- 平均パラメーターμと標準偏差パラメーターσの正規分布の数式表現 $p(y|\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(y-\mu)^2}{2\sigma^2}\}$
- > plot(y,dnorm(y,mean=0,sd=1),type="l")
- > pnorm(1.8,0,1)-pnorm(1.2,0,1)
- [1] 0.07913935
- > dnorm(1.5,0,1)*0.6
- [1] 0.07771056



6.8 ガンマ分布のGLM

- $p(y|s,r) = \frac{r^s}{\Gamma(s)} y^{s-1} \exp(-ry)$
- S→shapeパラメーター、r→rateパラメーター、$\Gamma(s)$→ガンマ関数
- Dgamma(y,shape,rate)関数で評価できます。
- 例題：50個体の葉の重量と花の重量の関係を調べます。i,xi,yi
- Yi→平均μiのガンマ分布
- >glm(y~log(x),family=Gamma(link="log"),data=d)