

3.6 説明変数が数量型 + 因子型の統計モデル

3.6 説明変数が数量型 + 因子型の統計モデル

説明変数：体サイズ \mathbf{x}_i と施肥効果 \mathbf{f}_i （重回帰）

$$\log \lambda_i = \beta_1 + \beta_2 \mathbf{x}_i + \beta_3 \mathbf{d}_i$$

ここまでのおさらい

リンク関数（対数）

種子数の平均

log(切片)

花 i のサイズ

log(\mathbf{f}_i の偏回帰係数)

花 i の施肥の有無（ \mathbf{f}_i を1, 0化したもの）

log(\mathbf{x}_i の偏回帰係数)

線形予測子

3.6 説明変数が数量型 + 因子型の統計モデル

実際にやってみる

```
> fit.all <- glm(y ~ x + f, data = d, family = poisson)
> fit.all

Call:  glm(formula = y ~ x + f, family = poisson, data = d)

Coefficients:
(Intercept)          x          fT
    1.26311    0.08007   -0.03200

Degrees of Freedom: 99 Total (i.e. Null);  97 Residual
Null Deviance:      89.51
Residual Deviance: 84.81      AIC: 476.6
```

切片が1.26、xの偏回帰係数が0.08、fがT（肥料あり）の偏回帰係数が-0.032？
つまり、サイズが1増えると種子が0.08個増えると解釈できる？
そして、肥料をあげると種子が0.032個減ると解釈できる？
・・・今回は対数リンク関数を用いているので違う

3.6 説明変数が数量型 + 因子型の統計モデル

Rで求めた回帰式

$$\text{肥料あり} : \log \lambda_i = 1.26 + 0.08x_i + 0.032$$

$$\text{肥料なし} : \log \lambda_i = 1.26 + 0.08x_i$$

λ を説明する式（肥料ありの場合）

$$\lambda_i = \exp(1.26) \times \exp(0.08x_i) \times \exp(-0.032)$$

→ x が1増えると λ が1.08倍になると解釈できる

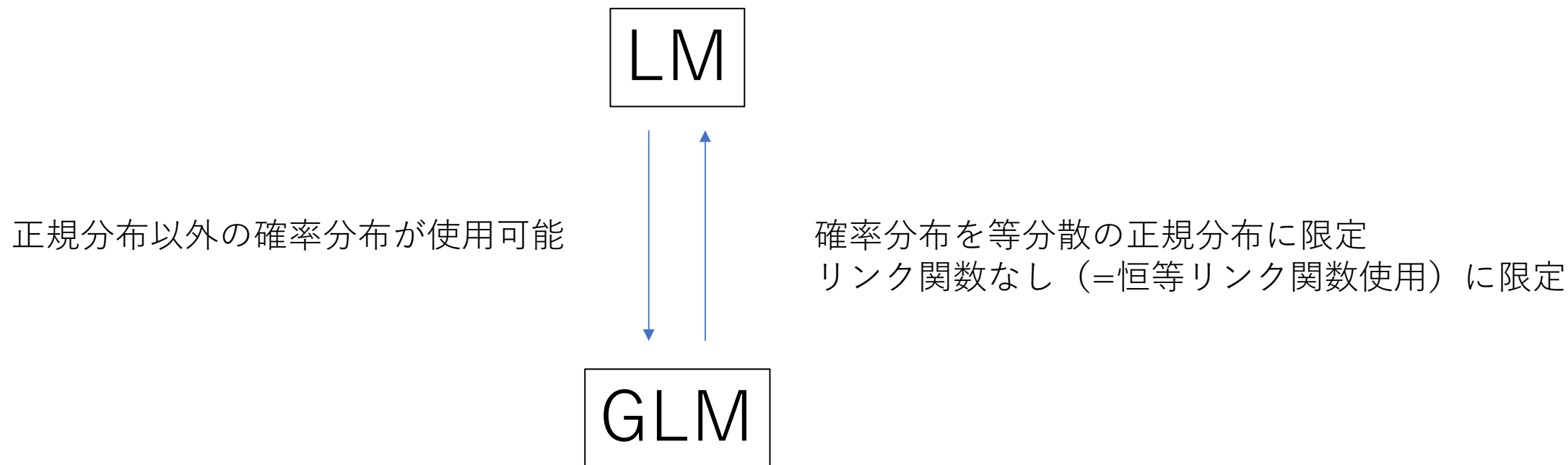
3.6 説明変数が数量型 + 因子型の統計モデル

リンク関数を用いない（=恒等リンク関数を用いる）
とどうなるのか？
続きはRで

3.7 「何でも正規分布」「何でも直線」には無理がある

3.7 「何でも正規分布」「何でも直線」には無理がある

一般化線形モデル（GLM）は線形モデル（LM）の拡張（復習ですが・・・）



3.7 「何でも正規分布」「何でも直線」には無理がある

正規分布だと無理がある場合の例

- ・ データが離散値である
 - この本では非推奨だが、正規分布で近似するという作法もあるらしい？

直線だと無理がある場合の例

- ・ カウントデータなのに平均値の予測がマイナスになる（p.61の図3.9参照）
 - ・ 直線回帰の当てはまりが悪い（分散が一定でないデータ、非線形データ）
- 今回カウントデータ（離散値、 >0 ）に対して対数リンク関数を使ったポアソン分布を用いたように、「データに適したモデリング」を行う

具体的な「データに適したモデリング」の方法についてはそのうちやと思うので、今回はここまで