

## 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

切片

花のサイズ

$\mathbf{d}_i$ の偏回帰係数

施肥の有無（ $\mathbf{f}_i$ を1, 0化したもの）

$\mathbf{x}_i$ 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増えると平均の種子数 $\lambda$ が0.08個増えると解釈できる？  
そして、肥料をあげると $\lambda$ 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

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

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

$\lambda$  を説明する式（肥料ありの場合）

$$\exp(0.08x_i) = e^{0.08x_i} \cong 1.08x_i$$

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

→  $x$ が1増えると  $\lambda$  が約1.08倍になると解釈できる

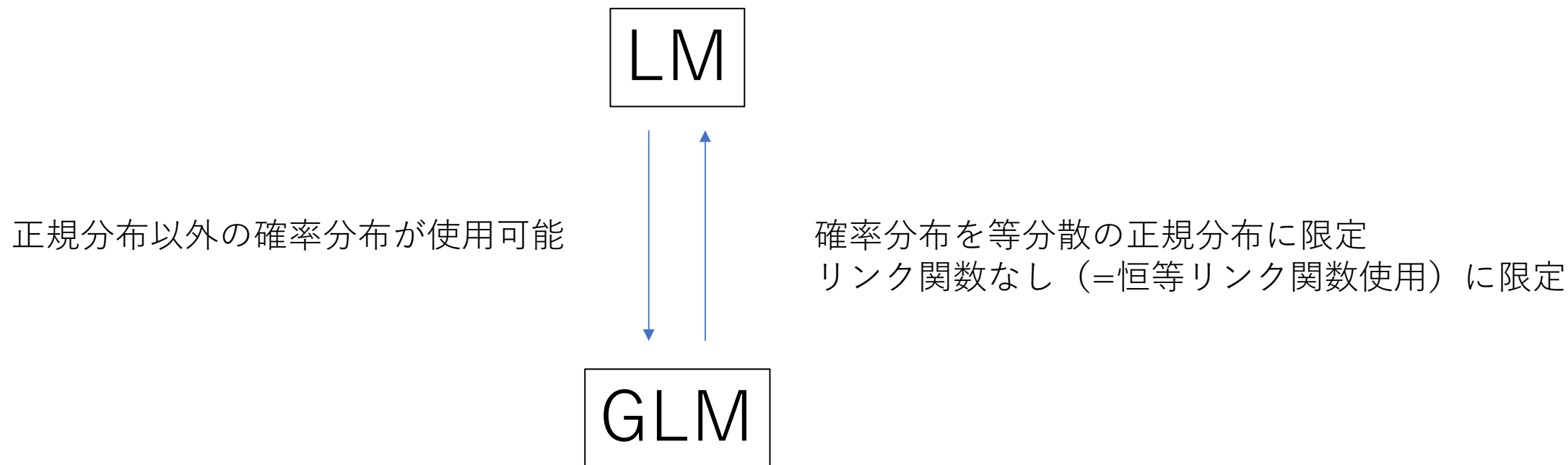
### 3.6 説明変数が数量型 + 因子型の統計モデル

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場合
  - 正規分布は連続値を扱う確率分布
  - 疑問「離散確率分布を正規分布で近似する」は無しなのか？(p.62 16行目)

#### 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場合 (p.61の図3.9)
- ・ 直線回帰の当てはまりが悪い場合 (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当てはめたように、「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