すべてのデータを有効利用しているとは言いがたく，またデータを取捨選択する手間もかかる．

そこで，モデルを考え直し，グループ差（会社差）に仮定を入れる．

8.1.4　階層モデル

- メカニズムの想像

各会社の $a[k]$ を「すべての会社で共通の全体平均」と「会社差を表す項」に分けて考える．そして，後者の会社差を表すパラメータは平均 0・標準偏差 σ_a の正規分布から生成されると考える．$b[k]$ についても同様である．このようにゆるい制約を入れて推定すると，例えば「『新卒の基本年収』の会社差のバラツキは σ_a ぐらいである」と言うことができる．また，例え数点しかデータがない会社があっても，そのデータを有効に利用して，その会社の a, b および全体の σ_a, σ_b の推定に活かすことができる．σ_a と σ_b の事前分布には無情報事前分布を設定してデータから推定する．階層的に事前分布を設定しているので**階層モデル** (hierarchical model) と呼ばれる．

- モデル式の記述

モデル式にすると以下になる．

■モデル式 8-3■

$$Y[n] \sim \mathrm{Normal}(a[KID[n]] + b[KID[n]]\ X[n], \sigma_Y) \qquad n = 1, \ldots, N \qquad (8.1)$$

$$a[k] = a_{全体平均} + a_{会社差}[k] \qquad k = 1, \ldots, K \qquad (8.2)$$

$$a_{会社差}[k] \sim \mathrm{Normal}(0, \sigma_a) \qquad k = 1, \ldots, K \qquad (8.3)$$

$$b[k] = b_{全体平均} + b_{会社差}[k] \qquad k = 1, \ldots, K \qquad (8.4)$$

$$b_{会社差}[k] \sim \mathrm{Normal}(0, \sigma_b) \qquad k = 1, \ldots, K \qquad (8.5)$$

データから $\sigma_Y, a_{全体平均}, a_{会社差}[k], \sigma_a, b_{全体平均}, b_{会社差}[k], \sigma_b$ を推定する．σ_a と σ_b だけでなく，$a_{全体平均}$ と $b_{全体平均}$ の事前分布にも無情報事前分布を設定する．モデル式 8-3 で，(8.2) 式を $a[k] = a_{全体平均}$ に，(8.4) 式を $b[k] = b_{全体平均}$ に変えるとモデル式 8-1 と一致する．また，モデル式 8-3 で，(8.3) 式と (8.5) 式がないとモデル式 8-2 と一致する．(8.3) 式や (8.5) 式で正規分布を使って $a_{会社差}[k]$ と $b_{会社差}[k]$ にゆるい制約を入れているのがポイントである．

- R でシミュレーション

モデル式 8-3 を使って，シミュレーションでデータを生成してみよう．σ_a と σ_b を無情報事前分布から乱数で生成してシミュレーションすると，最終的に得られる $Y_{sim}[n]$ は事前予測分布からの乱数サンプルに相当する（2.5 節参照）．しかし，それでは絶対値が非常に大きい $b[k]$ が頻繁に生成されるので，$Y_{sim}[n]$ はしばしば年収ではあり得ない値になってしまう．そこで，各パラメータに仮に定数を与え，シミュレーション結果を見てみよう．R コードの例は以下になる．

sim-model8-3.R

```
1  set.seed(123)
2  N <- 40
3  K <- 4
4  N_k <- c(15, 12, 10, 3)
5  a0 <- 350
6  b0 <- 12
7  s_a <- 60
8  s_b <- 4
9  s_Y <- 25
10 X <- sample(x=0:35, size=N, replace=TRUE)
11 KID <- rep(1:4, times=N_k)
12
13 a <- rnorm(K, mean=0, sd=s_a) + a0
14 b <- rnorm(K, mean=0, sd=s_b) + b0
15 d <- data.frame(X=X, KID=KID, a=a[KID], b=b[KID])
16 d <- transform(d, Y_sim=rnorm(N, mean=a + b*X, sd=s_Y))
```

2〜4 行目：それぞれの行で人数，会社の数，各会社に勤務している人数を与えている．

5〜9 行目：各パラメータに定数を与えている．モデル式 8-3 の $a_{全体平均}$, $b_{全体平均}, \sigma_a, \sigma_b, \sigma_Y$ がそれぞれ a0, b0, s_a, s_b, s_y にあたる．

13〜14 行目：4 社分の $a[k]$ と $b[k]$ を確率的に生成している．$a_{会社差}[k]$ が `rnorm(K, mean=0, sd=s_a)` に相当し，$b_{会社差}[k]$ が `rnorm(K, mean=0, sd=s_b)` に相当する．

16 行目：それらを使って y_{base} を計算し，さらにノイズ $\mathrm{N}(0, \sigma_Y)$ を加えて $Y_{sim}[n]$ としている．$Y_{sim}[n]$ の分布を確認するために散布図を描いた（図 8.2）．

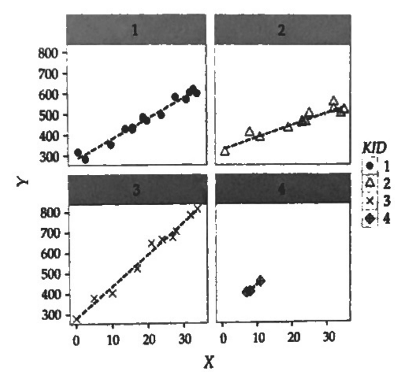

図 8.2　シミュレーションで生成したデータの例．凡例は図 8.1（右）と同じ．

sim-model8-3.R の 5〜9 行目で a0, b0, s_a, s_b, s_Y に対して定数を与えたが，これらの値をいろいろ変えたり，乱数の種を変えたりしながら図 8.2 の散布図を繰り返し描くと，このモデル式がどのようなデータを生成しやすいかを

視覚的に確認できる．モデル式が適切でないと，このシミュレーションの結果が意図しないものになる．また，シミュレーションで生成したデータに対して，後述の **model8-3.stan** を実行し，推定がうまくいくのか確認することも有益である．モデルが複雑になった場合，このようにRでシミュレーションをするのは大切なステップである．

- Stan で実装

  モデル式8-3の実装例は以下である．

model8-3.stan

```
1  data {
2    int N;
3    int K;
4    real X[N];
5    real Y[N];
6    int<lower=1, upper=K> KID[N];
7  }
8
9  parameters {
10   real a0;
11   real b0;
12   real ak[K];
13   real bk[K];
14   real<lower=0> s_a;
15   real<lower=0> s_b;
16   real<lower=0> s_Y;
17 }
18
19 transformed parameters {
20   real a[K];
21   real b[K];
22   for (k in 1:K) {
23     a[k] = a0 + ak[k];
24     b[k] = b0 + bk[k];
25   }
26 }
27
28 model {
29   for (k in 1:K) {
30     ak[k] ~ normal(0, s_a);
31     bk[k] ~ normal(0, s_b);
32   }
33
34   for (n in 1:N)
35     Y[n] ~ normal(a[KID[n]] + b[KID[n]]*X[n], s_Y);
36 }
```

model8-2.stan と比べると，a[k] を a0 と ak[k] から作っているので少し複雑になっている．

10〜16行目：モデル式8-3における $a_{全体平均}$, $b_{全体平均}$, $a_{会社差}[k]$, $b_{会社差}[k]$, σ_a, σ_b をそれぞれ a0, b0, ak, bk, s_a, s_b で宣言している．

20〜25行目：20〜21行目でモデル式8-3における $a[k]$ と $b[k]$ をそれぞれ a と b で宣言し，22〜25行目で定義している．

29〜32行目：$a_{会社差}[k]$ と $b_{会社差}[k]$ に対して正規分布を仮定し，ゆるい制約を入れている．

- 推定結果の解釈

  推定して得られた一部のパラメータの中央値と95%ベイズ信頼区間は以下の通りである．

$a_{全体平均}$：369.5 (179.2〜663.2)　$b_{全体平均}$：12.1 (−6.6〜29.0)
σ_a：94.8 (10.6〜684.0)　σ_b：8.2 (3.2〜42.7)　σ_Y：27.9 (22.2〜37.2)

「新卒の基本年収」の会社差のバラツキは94.8（万円）程度，「年齢に伴う昇給額」の会社差のバラツキは8.2（万円）程度，ノイズの大きさは27.9（万円）程度と解釈できる．また，同じように各会社の $a[k]$ と $b[k]$ についても中央値と区間を算出できる．

ここでは4社分のデータしかなかったので，各パラメータの95%ベイズ信頼区間は広く推定されている．例えば，$b_{全体平均}$ ではマイナスの値が，σ_a では600万円という大きな値が95%区間に含まれているが常識からは考えにくい．これに対処する方法の一つは弱情報事前分布を活用する方法であり，10.2節で扱う．

8.1.5 モデルの比較

ここで，モデル式8-1, モデル式8-2, モデル式8-3をグラフィカルモデルで比べてみよう．

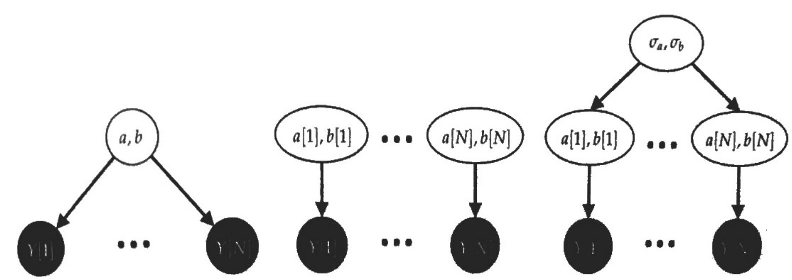

図8.3　各モデルのグラフィカルモデル．ただし，説明のために a と b をまとめたノードで表記している．（左）モデル式8-1,（中）モデル式8-2,（右）モデル式8-3.

図8.3において，Y は個人ごとのデータである．a と b はパラメータである．σ_a と σ_b は a と b を決めるハイパーパラメータである．このようにグラフィカルモデルで確認すると，モデルの違いが整理され，なぜモデル式8-3が階層モデル