

課題 11.E.4 略解例

表 2.E.2 のデータを、グループ変数 XA 、 XB と XC をダミー変数として設定するときの重回帰モデルの Stan スクリプトをリスト 11.E.4.1 のように用意した。

リスト 11.E.4.1 重回帰モデルの Stan スクリプト

```
data{
  int N;
  real Y[N];
  real X[N];
  real XA[N];
  real XB[N];
  real XC[N];
}
parameters{
  real b;
  real bA;
  real bB;
  real bC;
  real<lower = 0.0> sgm;
}
transformed parameters{
  real Yreg[N];
  for (i in 1:N)
    Yreg[i] = b * X[i] + bA * XA[i] + bB * XB[i] + bC * XC[i];
}
model{
  for (i in 1:N)
    Y[i] ~ normal(Yreg[i], sgm);
}
```

リスト 11.E.4.1 の Stan スクリプトを用いる Python スクリプトをリスト 11.E.4.2 のように用意した。

リスト 11.E.4.2 Stan スクリプト (リスト 11.E.4.1) を用いる Python スクリプト

```
import pystan
import numpy as np
import matplotlib.pyplot as plt

Data = [# ID SA SB
        ['A', 15, 15],
        ['A', 16, 16],
        ['A', 12, 19],
```

```

    ['A', 19, 12],
    ['B', 25, 25],
    ['B', 26, 26],
    ['B', 22, 29],
    ['B', 29, 22],
    ['C', 35, 35],
    ['C', 36, 36],
    ['C', 32, 39],
    ['C', 39, 32]
]

```

```
N = len(Data)
```

```
Y = []
```

```
X = []
```

```
XA = [0] * N
```

```
XB = [0] * N
```

```
XC = [0] * N
```

```
pos = 0
```

```
for v in Data:
```

```
    Y.append(v[2])
```

```
    X.append(v[1])
```

```
    if v[0] == 'A':
```

```
        XA[pos] = 1
```

```
    elif v[0] == 'B':
```

```
        XB[pos] = 1
```

```
    else:
```

```
        XC[pos] = 1
```

```
    pos += 1
```

```
DataStan = {'N': N, 'Y': Y, 'X': X, 'XA': XA, 'XB': XB, 'XC': XC}
```

```
sm = pystan.StanModel(file = 'MRegEx4.stan')
```

```
fit = sm.sampling(data = DataStan, n_jobs = 1)
```

```
print(fit)
```

```
smpl_b = fit['b']
```

```
n_smpls = len(smpl_b)
```

```
b_neg = 0
```

```
for v in smpl_b:
```

```
    if v < 0.0:
```

```
        b_neg += 1
```

```
plt.hist(smpl_b)
```

```
plt.xlabel('b')
```

```
plt.title('Prop(b < 0) = {0:.3f}'.format(b_neg / n_smpls))
```

```
plt.show()
```

```
smpl_Yregs = fit['Yreg']
```

```
Yreg = []
```

```
for i in range(N):
```

```

Yreg.append([])

for v in smpl_Yregs:
    pos = 0
    for u in v:
        Yreg[pos].append(u)
        pos += 1

meanYreg = [0] * N
for i in range(N):
    meanYreg[i] = np.median(Yreg[i])

plt.plot(meanYreg, Y, 'bo')
plt.xlabel('Yreg')
plt.ylabel('Y')
minYreg = np.amin(meanYreg)
maxYreg = np.amax(meanYreg)
plt.plot([minYreg, maxYreg], [minYreg, maxYreg], 'b--', label = 'y = x')
plt.legend()
plt.show()

```

リスト 11.E.4.2 のスクリプトを実行すると、図 11.E.4.1 のフォームが表示される。

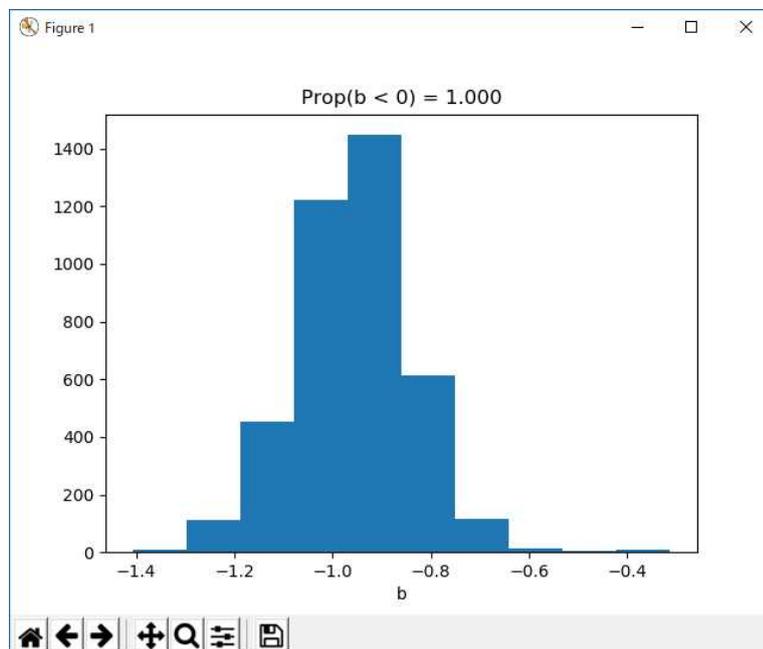


図 11.E.4.1 傾き b の事後分布

傾き b の事後分布における b の値が負である比率は「 $\text{Prop}(b < 0) = 1.000$ 」であり、0 の負の側に事後分布が位置しており、傾き b は負であると考えられる。単回帰モデルによる

分析では、正であった (図 11.E.2.1)。

図 11.E.4.1 のフォームを閉じるとお、図 11.E.4.2 のフォームが表示される。

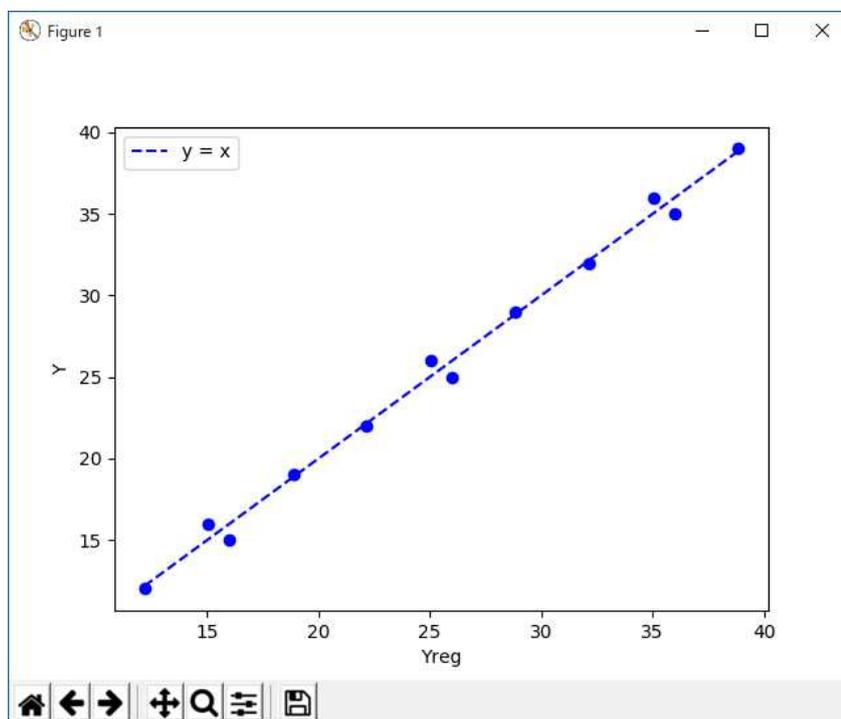


図 11.E.4.2 予測値の事後分布の中央値 Yreg とデータ Y。
予測値とデータが一致すると点(Yreg, Y)は直線 $y = x$ 上にある。

図 11.E.4.2 のグラフは、予測値 (事後分布の中央値) とデータ値の関係を表すものである。予測値とデータ値が一致すれば、直線 $y = x$ 上に点 (予測値、データ値) がプロットされる。直線 $y = x$ と点の散布図との関係に特に偏りは認められないので、モデルとデータとの関係に特に問題はないと考えられる。

独立変数 SA の偏回帰係数 b が、図 11.E.4.1 に示されている事後分布から負であると考えられ、さらに、モデルとデータとの関係も図 11.E.4.2 に示されている関係から特に問題はない考えられるが、これはグループの要因を表すダミー変数 XA、XB と XC を導入すると、単回帰分析 (課題 11.E.2) での変数 SA と従属変数 SB の間の 1 次式の関係、傾きが正、は逆転して負になることを示している。