

課題 11.E.1 略解

Stan スクリプトをリスト 11.E.1.1 のように作成した。

リスト 11.E.1.1 単回帰モデルの Stan スクリプト

```
data{
  int N;
  real Y[N];
  real X[N];
}
parameters{
  real b0;
  real b1;
  real<lower = 0.0> sgm;
}
transformed parameters{
  real Yreg[N];
  for (i in 1:N)
    Yreg[i] = b0 + b1 * X[i];
}
model{
  for (i in 1:N)
    Y[i] ~ normal(Yreg[i], sgm);
}
```

リスト 11.E.1.1 の Stan スクリプトを用いた単回帰分析の Python スクリプトをリスト 11.E.1.2 のように用意した。

リスト 11.E.1.2 リスト 11.E.1.1 の Stan スクリプトを用いた Python スクリプト

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

```
Data = [# ID SA SB
        ['A', 11, 10],
        ['A', 12, 11],
        ['A', 11, 12],
        ['A', 10, 11],
        ['A', 11, 11],
        ['B', 13, 12],
        ['B', 14, 13],
        ['B', 13, 14],
        ['B', 12, 13],
        ['B', 13, 13],
        ['C', 15, 14],
```

```

        ['C', 16, 15],
        ['C', 15, 16],
        ['C', 14, 15],
        ['C', 15, 15]
    ]

N = len(Data)
Y = []
X = []
for v in Data:
    Y.append(v[2])
    X.append(v[1])

DataStan = {'N': N, 'Y': Y, 'X': X}

sm = pystan.StanModel(file = 'SRegEx.stan')

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

smp1_b1 = fit['b1']
n_smp1s = len(smp1_b1)
b1_pos = 0
for v in smp1_b1:
    if v > 0.0:
        b1_pos += 1
plt.hist(smp1_b1)
plt.xlabel('b1')
plt.title('Prop(b1 > 0) = {0:.3f}'.format(b1_pos/n_smp1s))
plt.show()

smp1_Yregs = fit['Yreg']

Yreg = []
for i in range(N):
    Yreg.append([])

for v in smp1_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])
minYreg = np.amin(meanYreg)
maxYreg = np.amax(meanYreg)
plt.plot([minYreg, maxYreg], [minYreg, maxYreg], 'b--', label = 'y = x')
plt.plot(meanYreg, Y, 'bo')
plt.xlabel('Yreg')
plt.ylabel('Y')
plt.legend()
plt.show()

```

リスト 11.E.1.2 のスクリプトを実行すると、図 11.E.1.1 のフォームが表示される。

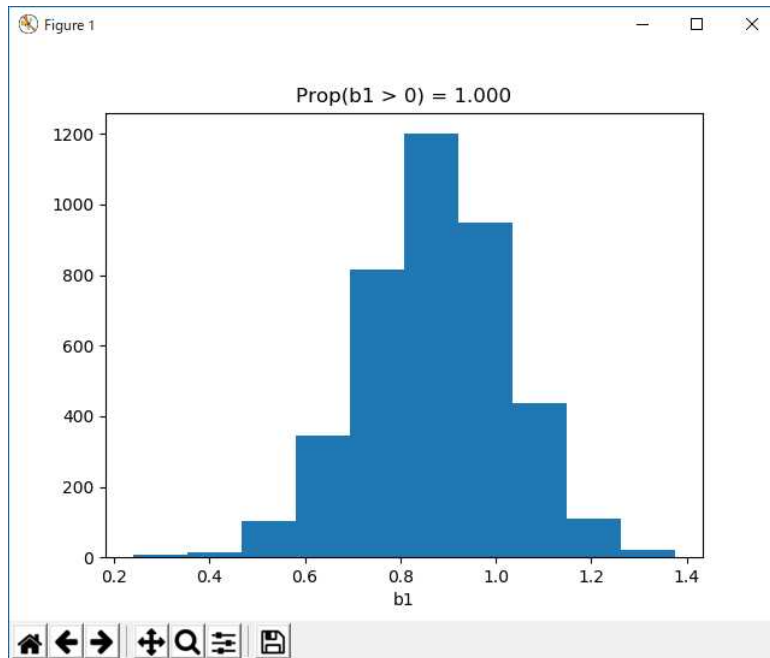


図 11.E.1.1 リスト 11.E.1.2 の実行時に表示されるグラフ

グラフ上部に「Prop($b_1 > 0$) = 1.000」とあるが、これは単回帰モデルにおける 1 次関数の傾き b_1 の事後分布からのサンプリングにおいて、その値が正であったものの比率を示すものである。傾き b_1 の値が正であった比率が 1.000 であるので、有意に傾きが正であると言える。

図 11.E.1.1 のフォームを閉じると、図 11.E.1.2 のフォームが表示される。

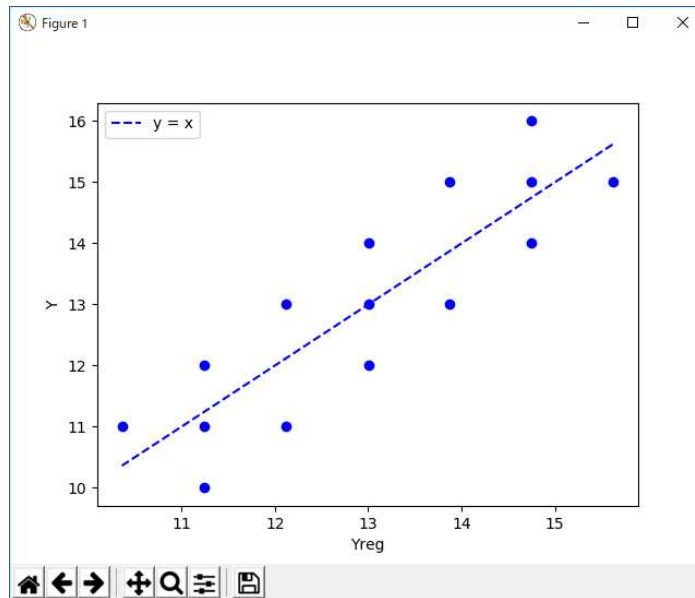


図 11.E.1.2 単回帰モデルによる予測値 Yreg とデータ値 Y の関係

横軸に単回帰モデルによる予測値、縦軸にデータ値をとったものである。予測値は、事後分布における中央値である。予測値とデータ値が一致すれば、直線 $y=x$ 上に点がプロットされる。

図 11.E.1.2 のフォームを閉じると、プログラムの実行が終了する。