

課題 11.E.4 略解例

表 2.E.2 のデータを、グループ変数 X_A 、 X_B と X_C をダミー変数として設定するときの重回帰モデルの Python スクリプトをリスト 14.E.4.1 のように用意した。

リスト 14.E.4.1 重回帰モデルの Python スクリプト

```
import pymc3 as pm
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

with pm.Model() as MReg4:
    b = pm.Uniform('b', lower = -1000, upper = 1000)
    bA = pm.Uniform('bA', lower = -1000, upper = 1000)
    bB = pm.Uniform('bB', lower = -1000, upper = 1000)
    bC = pm.Uniform('bC', lower = -1000, upper = 1000)
    Y_est = b * X + bA * XA + bB * XB + bC * XC
    sgm = pm.Uniform('sgm', lower = 0, upper = 1000)
```

```

Y_obs = pm.Normal('Y_obs', mu = Y_est, sd = sgm, observed = Y)
trace = pm.sample()
pm.traceplot(trace)
plt.show()

summary_trace = pm.summary(trace)
print('\n', summary_trace)

smp1_b = trace['b']

bp025 = np.percentile(smp1_b, 2.5)
bp975 = np.percentile(smp1_b, 97.5)
plt.hist(smp1_b)
plt.xlabel('b')
plt.title('95% CI = [{0:.3f}, {1:.3f}]'.format(bp025, bp975))
plt.show()

smp1_bA = trace['bA']
smp1_bB = trace['bB']
smp1_bC = trace['bC']

Yregs = []
for j in range(N):
    Yregs.append([])

for sb, sbA, sbB, sbC in zip(smp1_b, smp1_bA, smp1_bB, smp1_bC):
    for j in range(N):
        v = sb * X[j] + sbA * XA[j] + sbB * XB[j] + sbC * XC[j]
        Yregs[j].append(v)

meanYreg = [0] * N
for i in range(N):
    meanYreg[i] = np.median(Yregs[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 14.E.4.1 のスクリプトを実行すると、図 14.E.4.1 のフォームが表示される。

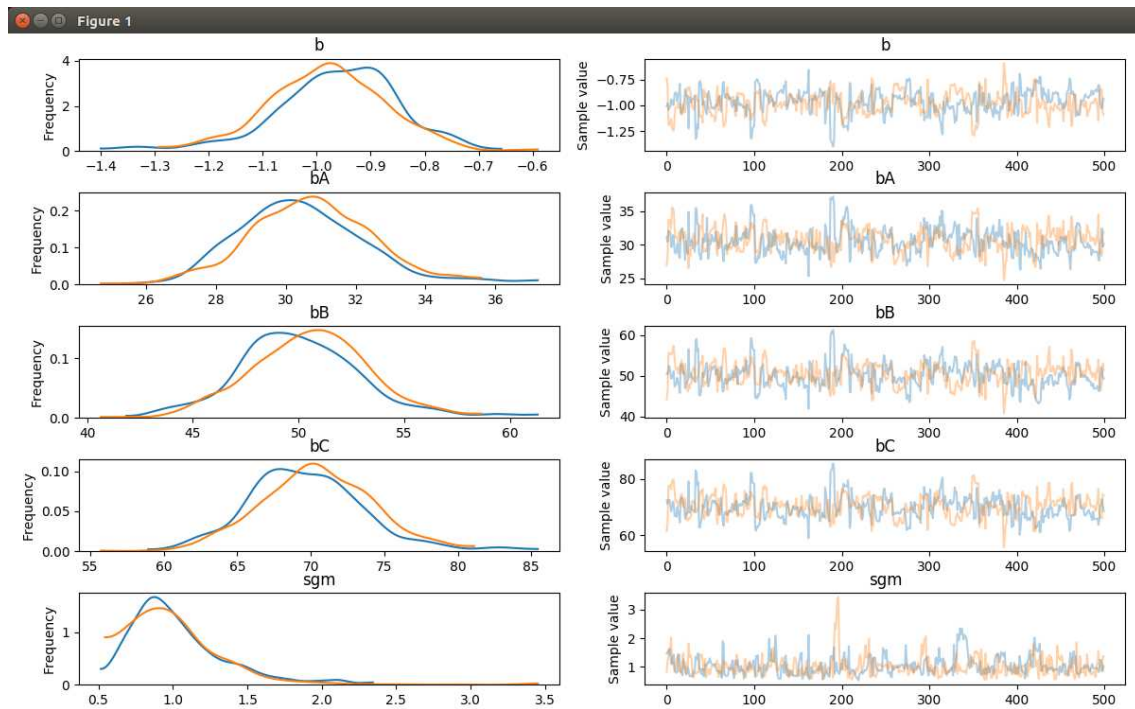


図 14.E.4.1 リスト 14.E.4.1 のスクリプトの実行で表示されるフォーム

図 14.E.4.1 のフォームを閉じると、図 14.E.4.2 のフォームが表示される。

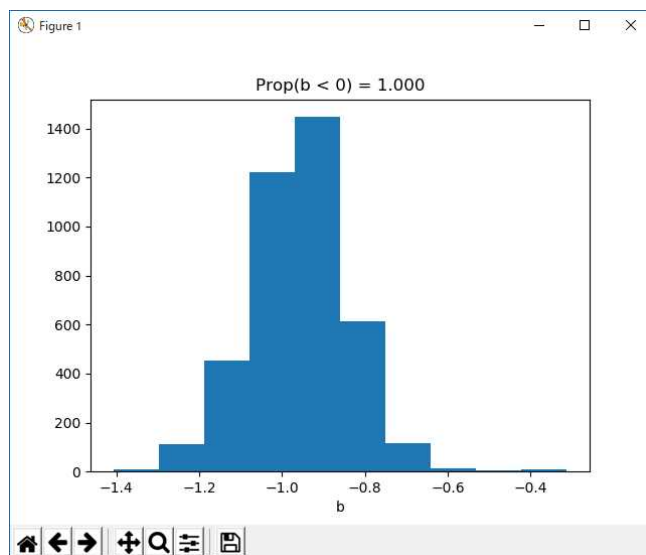


図 14.E.4.2 傾き b の事後分布

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 14.E.2.2)。

図 14.E.4.2 のフォームを閉じるとお、図 14.E.4.3 のフォームが表示される。

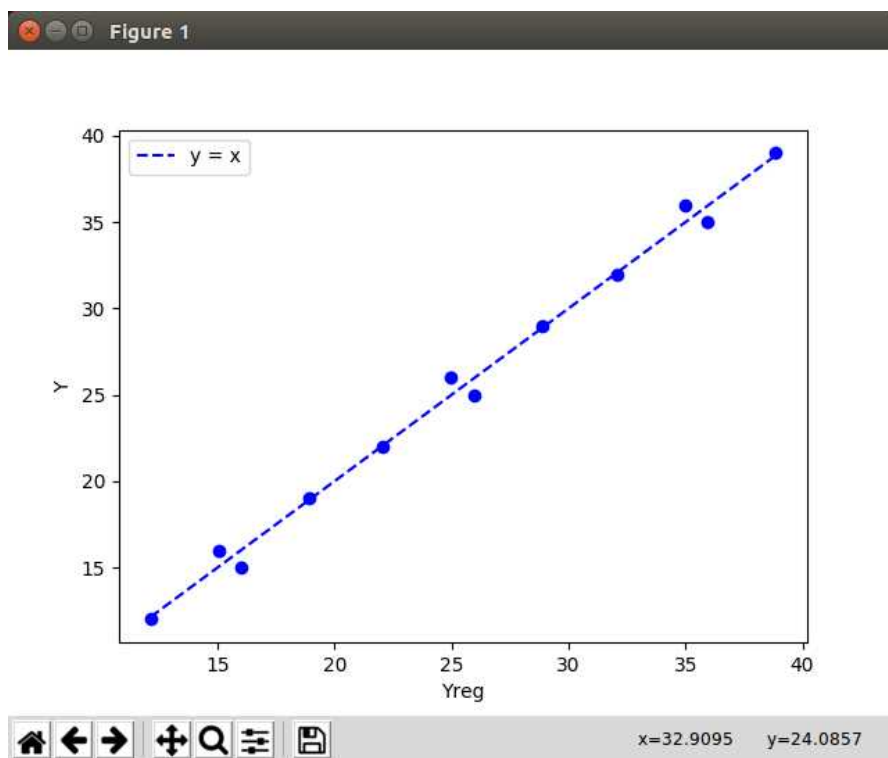


図 14.E.4.3 予測値の事後分布の中央値 Y_{reg} とデータ Y 。
予測値とデータが一致すると点 (Y_{reg}, Y) は直線 $y = x$ 上にある。

図 14.E.4.3 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 14.E.4.2 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 14.E.4.3 に示されている関係から特に問題はないと考えられるが、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 14.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。