

データサイエンス 課題3

2323050 井上祐斗

問1

A党候補者の支出(EXP_A)とB党候補者の支出(EXP_B)が、A党候補者の得票率(PCT_Vote_A)に与える影響を調べるために、重回帰分析を実行しました。

```
# データの読み込み
vote2 <- read.csv("vote2.csv")

# 重回帰分析の実行
model1 <- lm(PCT_Vote_A ~ EXP_A + EXP_B, data = vote2)

summary(model1)
```

```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 49.561548  1.433270  34.58  <2e-16 ***
EXP_A       0.038268  0.003396  11.27  <2e-16 ***
EXP_B      -0.035992  0.003124 -11.52  <2e-16 ***
```

解釈

- 切片 (Intercept): A党とB党の選挙費用がどちらも0の場合、A党の得票率は約49.6%になると予測されます。
- EXP_A の係数: B党の選挙費用が一定の場合、A党の選挙費用が1000ドル増加すると、A党の得票率は約0.038%増加することを示します。
- EXP_B の係数: A党の選挙費用が一定の場合、B党の選挙費用が1000ドル増加すると、A党の得票率は約0.036%減少することを示します。

問2

次に、A党候補者の支出の効果が逡減するかどうかを調べるため、EXP_Aの2乗の項を追加した重回帰分析を行いました。

```
# 2乗項を含む重回帰分析の実行
model2 <- lm(PCT_Vote_A ~ EXP_A + I(EXP_A^2) + EXP_B, data = vote2)
```

```
summary(model2)
```

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.349e+01 1.648e+00 26.396 < 2e-16 ***

EXP_A 7.562e-02 6.924e-03 10.921 < 2e-16 ***

I(EXP_A^2) -3.944e-05 6.544e-06 -6.027 1.03e-08 ***

EXP_B -3.145e-02 2.939e-03 -10.701 < 2e-16 ***

解釈

EXP_Aの2乗の係数が負の値(-3.944e-05)で、統計的に有意($p < 0.001$)であるため、A党候補者の支出の効果が逡減することを示す証拠があります。

- \$500,000 の場合: 傾きは約 0.036 です。
- \$700,000 の場合: 傾きは約 0.020 です。

これは、選挙費用が増えるにつれて、追加の支出がもたらす得票率の増加が小さくなることを意味します。

問3

睡眠時間(sleep)を予測するモデルを、労働時間(totwrk)、教育年数(educ)、性別(male)、婚姻状況(marr)を用いて記述します。

$$\text{sleep} = \beta_0 + \beta_1 * \text{totwrk} + \beta_2 * \text{educ} + \beta_3 * \text{male} + \beta_4 * \text{marr} + u$$

- 基本グループ: maleとmarrが0であることを示す未婚女性が基準となる。

未婚女性	$E(\text{sleep} \mid \text{totwrk}, \text{educ}, \text{male}=0, \text{marr}=0) = \beta_0 + \beta_1 * \text{totwrk} + \beta_2 * \text{educ}$
既婚女性	$E(\text{sleep} \mid \text{totwrk}, \text{educ}, \text{male}=0, \text{marr}=1) = \beta_0 + \beta_1 * \text{totwrk} + \beta_2 * \text{educ} + \beta_4$
未婚男性	$E(\text{sleep} \mid \text{totwrk}, \text{educ}, \text{male}=1, \text{marr}=0) = \beta_0 + \beta_1 * \text{totwrk} + \beta_2 * \text{educ} + \beta_3$
既婚男性	$E(\text{sleep} \mid \text{totwrk}, \text{educ}, \text{male}=1, \text{marr}=1) = \beta_0 + \beta_1 * \text{totwrk} + \beta_2 * \text{educ} + \beta_3 + \beta_4$

問4

問3のモデルをRで推定し、結果を解釈します。

```
# データの読み込み
```

```
sleep2 <- read.csv("sleep2.csv")
```

```
# 重回帰分析の実行
```

```
model3 <- lm(sleep ~ totwrk + educ + male + marr, data = sleep2)
```

```
summary(model3)
```

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3720.94016 89.15542 41.735 <2e-16 ***

totwrk -0.16612 0.01806 -9.200 <2e-16 ***

educ -13.73009 5.66796 -2.422 0.0157 *

male 86.14584 34.96080 2.464 0.0140 *

marr 30.10525 41.94060 0.718 0.4731

解釈

- totwrk: 労働時間が1分増えると、睡眠時間は約0.17分減少します。この変数は統計的に有意です。
- educ: 教育年数が1年増えると、睡眠時間は約13.7分減少します。この変数は統計的に有意です。
- male: 男性は女性よりも睡眠時間が約86.1分長いです。この変数は統計的に有意です。
- marr: 婚姻状況は、このモデルでは睡眠時間に対して統計的に有意な影響を与えていません。