

# 計量経済学応用

## 課題 4 の解答例

矢内 勇生

2018-05-24

### 準備

まず、今回使うパッケージを読み込む。インストール済みでない場合は、`install.packages()` でインストールしてから使う。

```
library("foreign")
library("dplyr")
library("arm")
library("coefplot")
library("interplot")
library("texreg")
# Mac ユーザは次の行も実行
##theme_set(theme_gray(base_size = 12, base_family = "HiraKakuProN-W3"))
```

データ (hr96-09.dta) をダウンロードし、プロジェクト内の data フォルダに保存する。保存ができたならデータを読み込んで中身を確認する。

```
HR <- read.dta("data/hr96-09.dta")
glimpse(HR) # 結果は省略
```

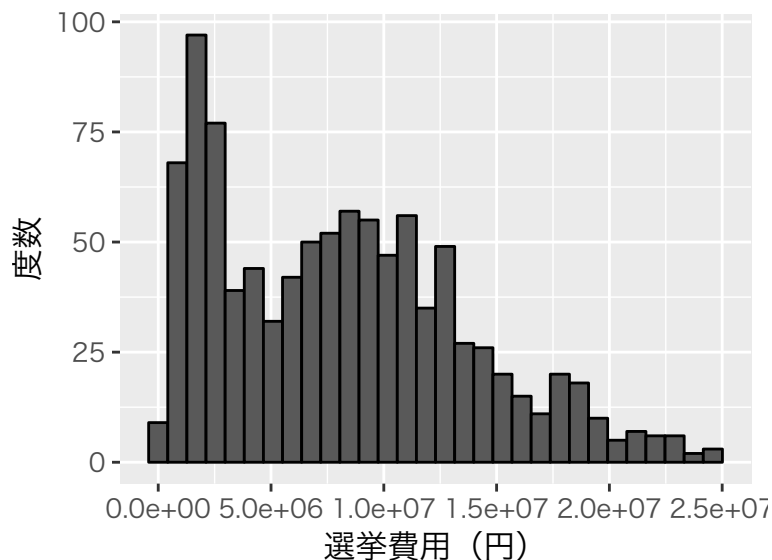
今回の課題では 2005 年の選挙が分析の対象なので、2005 年の分だけ取り出す。また、欠測値がある観測を除外する。

```
HR_2005 <- HR %>%
  filter(year == 2005) %>%
  na.omit()
```

### 1 投票率を選挙費用に回帰する

投票率 (voteshare) を選挙費用 (exp) に回帰する。回帰する前に、説明変数の分布を確かめよう。

```
exp_hist <- ggplot(HR_2005, aes(x = exp)) +
  geom_histogram(color = "black") +
  labs(x = "選挙費用 (円)", y = "度数")
print(exp_hist)
```



この `exp` という変数は、選挙費用を 1 円単位で測定している。1 円が得票率に与える影響は非常に小さいことが想定されるので、`expm` を 10 万で割って、選挙費用を 10 万円単位で測定する `exp2` という変数を作ろう。さらに、その変数を標本平均で中心化した `c_exp2` という変数を作り、得票率をその変数に回帰しよう。

```
HR_2005 <- HR_2005 %>%
  mutate(exp2 = exp / 10^5) %>%
  mutate(c_exp2 = exp2 - mean(exp2))
```

回帰係数を推定する。

```
fit_1 <- lm(voteshare ~ c_exp2, data = HR_2005)
display(fit_1, detail = TRUE)

## lm(formula = voteshare ~ c_exp2, data = HR_2005)
##               coef.est coef.se t value Pr(>|t|)
## (Intercept) 30.35      0.40  75.77   0.00
## c_exp2       0.26      0.01  36.23   0.00
## ---
## n = 985, k = 2
## residual sd = 12.57, R-Squared = 0.57
```

この結果を図示する。

```
plt_fit_1 <- ggplot(HR_2005, aes(x = c_exp2, y = voteshare)) +
  geom_smooth(method = "lm") +
  geom_point() +
  labs(x = "選挙費用の偏差 (10 万円)", y = "得票率 (%)")
print(plt_fit_1)
```

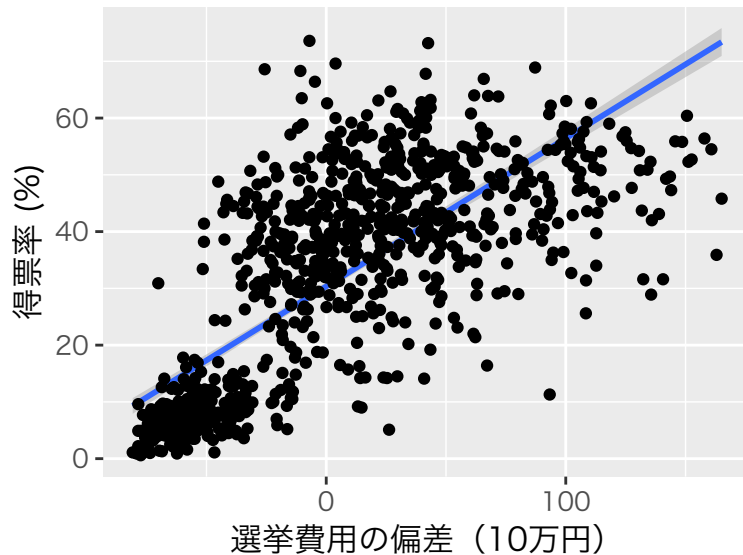


表 1 (最後のページ) のモデル 1 の列を見ると、選挙費用の回帰係数は 0.26 である。この回帰係数の  $p$  値は 0.05 より小さいので、有意水準 5 % で、選挙費用が得票率に与える影響は統計的に有意である。選挙費用を 10 万円増やすごとに、得票率が 0.26 ポイント上昇することが期待される。100 万円増やせば、得票率が 2.6 ポイント上昇することが期待される。選挙費用の分布を見てみると、候補者によって選挙費用が 100 万円以上異なるということはよくあるようである。つまり、選挙費用が 100 万円以上変化することは、現実として起こり得る。また、得票率が 2.6 ポイント変化するというのは、実際の選挙においては勝敗を左右し得る大きな変化であると考えられる。したがって、選挙費用が得票率に与える影響は、実質的にも有意であると考えられる。

## 2. 得票率を選挙費用と自民党ダミーに回帰する

まず、政党を表す変数 `party` を元に、自民党 (Liberal Democratic Party: LDP) を表すダミー変数 `ldp` を作る。

```
HR_2005 <- HR_2005 %>%
  mutate(ldp = as.numeric(party == "LDP"))
```

思い通りのダミー変数ができたかどうか確認する。

```
with(HR_2005, table(party, ldp))
```

```
##      ldp
## party  0  1
## msz    67  0
## JCP   275  0
## LDP     0 289
## CGP     9  0
## oki     0  0
## tai     0  0
## saki     0  0
## NFP     0  0
## DPJ   289  0
```

```
##   SDP    38   0
##   LF     0   0
##   NJSP   0   0
##   DRF    0   0
##   kobe   0   0
##   nii    0   0
##   sei     0   0
##   JNFP   0   0
##   bunka  0   0
##   green  0   0
##   LP     0   0
##   RC     0   0
##   muk    0   0
##   CP     0   0
##   NCP    0   0
##   ND     0   0
##   son    0   0
##   sek    1   0
##   NP     0   0
##   NNP    10   0
##   NPJ    6   0
##   NPD    1   0
##   minna  0   0
##   R      0   0
##   H      0   0
##   sho    0   0
```

ldp の値を見てみると、party の値が “LDP” のときは 1、それ以外の値のときは 0 になっていることが確認できる。

この変数を使って、重回帰分析を行う。推定するモデルは、

$$\text{得票率} = \beta_0 + \beta_1 \text{選挙費用(10万円)} + \beta_2 \text{自民党ダミー} + \text{誤差}$$

である。

```
fit_2 <- lm(voteshare ~ c_exp2 + ldp, data = HR_2005)
display(fit_2, detail = TRUE)
```

```
## lm(formula = voteshare ~ c_exp2 + ldp, data = HR_2005)
##               coef.est coef.se t value Pr(>|t|)
## (Intercept)  26.23      0.45   57.81   0.00
## c_exp2        0.20      0.01   26.14   0.00
## ldp          14.06      0.93   15.04   0.00
## ---
## n = 985, k = 3
## residual sd = 11.34, R-Squared = 0.65
```

この結果を図示する。まず、予測値のデータフレームを作る。

```
pred <- with(HR_2005, expand.grid(
  c_exp2 = seq(min(c_exp2), max(c_exp2), length.out = 100),
  ldp = c(0,1)
))
```

次に、説明変数に特定の値を代入した場合の予測値を計算する。

```
pred <- mutate(pred, voteshare = predict(fit_2, newdata = pred))
```

続いて、信頼区間の下限と上限を計算するために、予測値と標準誤差を求める。

```
err <- predict(fit_2, newdata = pred, se.fit = TRUE)
```

この予測値と標準誤差を使って 95% 信頼区間を求める。

```
pred$lower <- err$fit + qt(0.025, df = err$df) * err$se.fit
pred$upper <- err$fit + qt(0.975, df = err$df) * err$se.fit
```

以上の結果を使って図を作る。

```
## 散布図 (point) は元データ (HR_2005) で描き、回帰直線は予測値 (pred) で描く
plt_fit_2 <- ggplot(HR_2005, aes(x = c_exp2, y = voteshare, color = as.factor(ldp))) +
  geom_smooth(data = pred, aes(ymin = lower, ymax = upper), stat = 'identity') +
  geom_point(size = 1) +
  geom_line(data = pred) +
  labs(x = '選挙費用の偏差 (10 万円)', y = '得票率 (%)') +
  scale_color_discrete(name = '政党', labels = c('自民党以外', '自民党')) +
  guides(color = guide_legend(reverse = TRUE))
print(plt_fit_2)
```

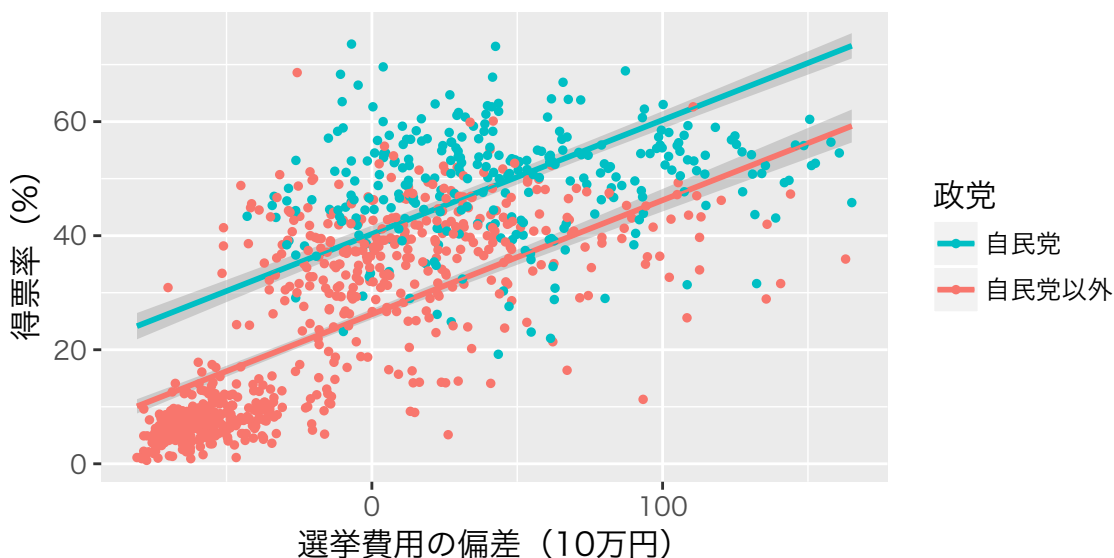


表 1 (最後のページ) のモデル 2 の列を見ると、選挙費用の回帰係数は 0.20 である。この回帰係数の  $p$  値は 0.05 より小さいので、有意水準 5 % で、選挙費用が得票率に与える影響は統計的に有意である。選挙費用を 100 万 (10 万) 円増やすごとに、得票率が 2 (0.2) ポイント上昇することが期待される。すでに述べたとおり、

選挙費用の分布を見てみると、候補者によって選挙費用が100万円以上異なるということはよくあるようである。つまり、選挙費用が100万円以上変化することは、現実として起こり得る。また、得票率が2ポイント変化するというのは、実際の選挙においては勝敗を左右し得る大きな変化であると考えられる。したがって、選挙費用が得票率に与える影響は、実質的にも有意であると考えられる。また、自民党ダミーの係数は14.06である。この推定値の  $p$  値も0.05より小さいので、有意水準5%で統計的に有意である。選挙費用の額が同じ候補者同士を比べると、自民党候補の方がそれ以外の候補よりも14ポイント以上高い得票が見込めるということを示しており、この効果は実質的にも大きいと言える。

### 3. 選挙費用と自民党ダミーの相互作用を考える

ここでは、以下の回帰モデルを考える。

$$\text{得票率} = \beta_0 + \beta_1 \text{選挙費用(10万円)} + \beta_2 \text{自民党ダミー} + \beta_3 \text{選挙費用(10万円)} \times \text{自民党ダミー} + \text{誤差}.$$

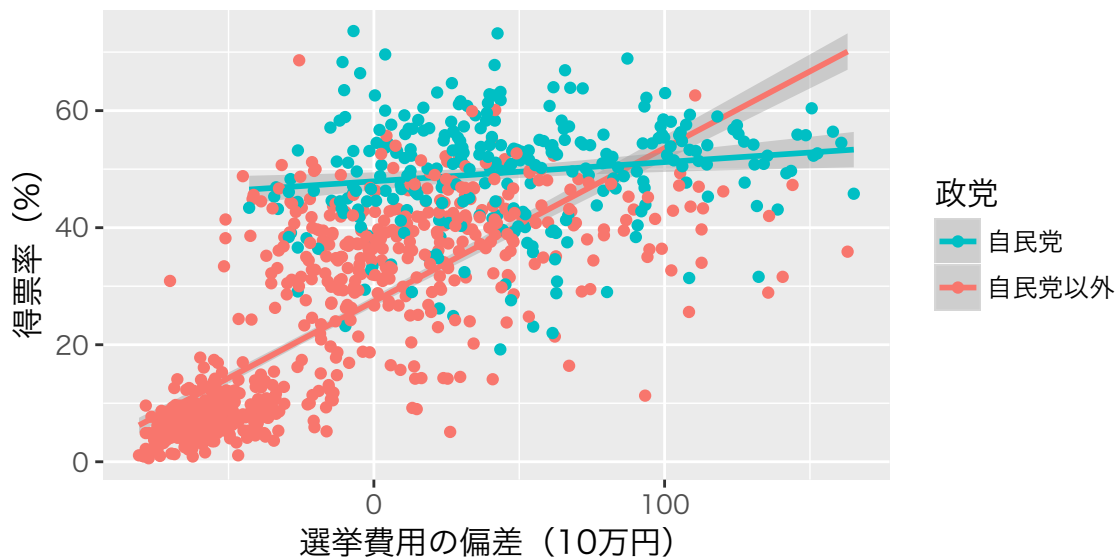
Rで回帰係数を推定する。

```
fit_3 <- lm(voteshare ~ c_exp2 * ldp, data = HR_2005)
display(fit_3, detail = TRUE)

## lm(formula = voteshare ~ c_exp2 * ldp, data = HR_2005)
##               coef.est coef.se t value Pr(>|t|)
## (Intercept)   27.40      0.42   65.47    0.00
## c_exp2         0.26      0.01   32.31    0.00
## ldp           20.52      0.95   21.52    0.00
## c_exp2:ldp    -0.23      0.02  -14.71    0.00
## ---
## n = 985, k = 4
## residual sd = 10.27, R-Squared = 0.71
```

この結果を図示する。

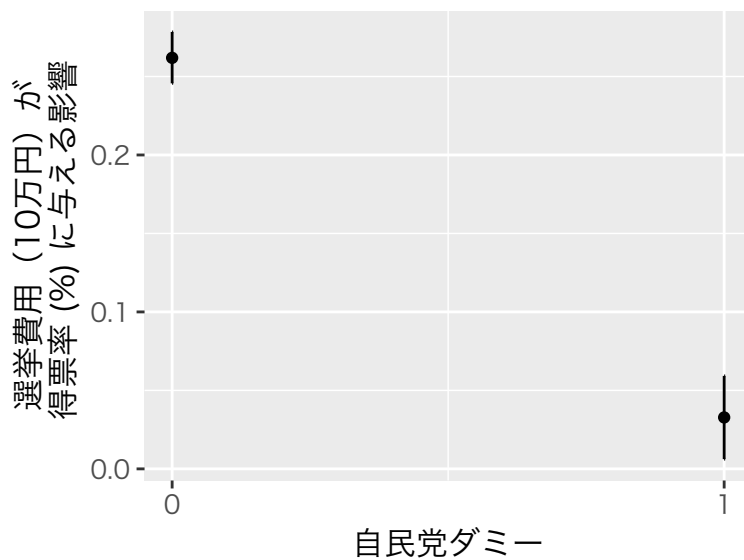
```
plt_fit_3 <- ggplot(HR_2005, aes(x = c_exp2, y = voteshare, color = as.factor(ldp))) +
  geom_smooth(method = "lm") +
  geom_point() +
  labs(x = '選挙費用の偏差 (10 万円)', y = '得票率 (%)') +
  scale_color_discrete(name = '政党', labels = c('自民党以外', '自民党')) +
  guides(color = guide_legend(reverse = TRUE))
print(plt_fit_3)
```



自民党以外の候補者については、選挙費用が増えるほど得票率が上がるという傾向がある。しかし、自民党の候補者については、選挙費用が増えても、得票率はほとんど変化しないようである。

表 1 (最後のページ) のモデル 3 の列を見ると、すべての説明変数について  $p$  値が 0.05 より小さくなっている。したがって、一見すると有意水準 5 % で、選挙費用、自民党ダミー、選挙費用と自民党ダミーの交差項のそれぞれが得票率に与える影響が統計的に有意であるように見えるが、本当にそうだろうか。

```
interplot(fit_3, var1 = "c_exp2", var2 = "ldp") +
  labs(x = "自民党ダミー", y = "選挙費用 (10 万円) が \n 得票率 (%) に与える影響")
```



この図では、点が係数の点推定値、線分が係数の 95 % 信頼区間を示している。自民党候補者にとっても民主党候補者にとっても、95% 信頼区間が 0 に重なっておらず、選挙費用の効果が 5% 水準で統計的に有意であることがわかる。また、選挙費用の影響の大きさについては、自民党候補者とそれ以外の候補者の間に違いがある (2 つの信頼区間は重ならず、区間がはっきり区別できる) ことがわかる。

表 1 (最後のページ) のモデル 3 の列から効果量 (係数の推定値) を読むと、自民党以外の候補者については、選挙費用の影響は 0.26 であることがわかる。つまり、選挙費用を 100 万円増やすごとに、平均すると得票率が 2.6 ポイント増加することがわかる。この効果は、実質的にも有意な効果であると考えられる。他方、自民

党候補の場合についてみると、選挙費用の影響は  $0.26 - 0.23 = 0.03$  である。つまり、自民党候補者の場合、選挙費用を 10 万円増やすごとに得票率が 0.03 ポイント上昇することがわかる。あるいは、選挙費用を 500 万円増やすごとに、得票率が 1.5 ポイント上がる。選挙費用を 500 万円増やすのが困難であり、1.5 ポイント得票率がそれほど結果に影響しないと考えれば、この効果は実質的にはあまり大きな意味を持たないだろう。

## 4. 得票率を選挙費用、有権者数と 2 変数の相互作用に回帰する

ここでは、以下の回帰モデルを考える。

$$\text{得票率} = \beta_0 + \beta_1 \text{選挙費用(10万円)} + \beta_2 \text{有権者数} + \beta_3 \text{選挙費用(10万円)} \times \text{有権者数} + \text{誤差}.$$

R で回帰係数を推定する。ただし、有権者数は、測定単位を 1 人から 1 万に変換し、標本平均で中心化する。

```
HR_2005 <- mutate(HR_2005, c_eligible = (eligible - mean(eligible)) / 10000)
fit_4 <- lm(voteshare ~ c_exp2 * c_eligible, data = HR_2005)
display(fit_4, detail = TRUE)
```

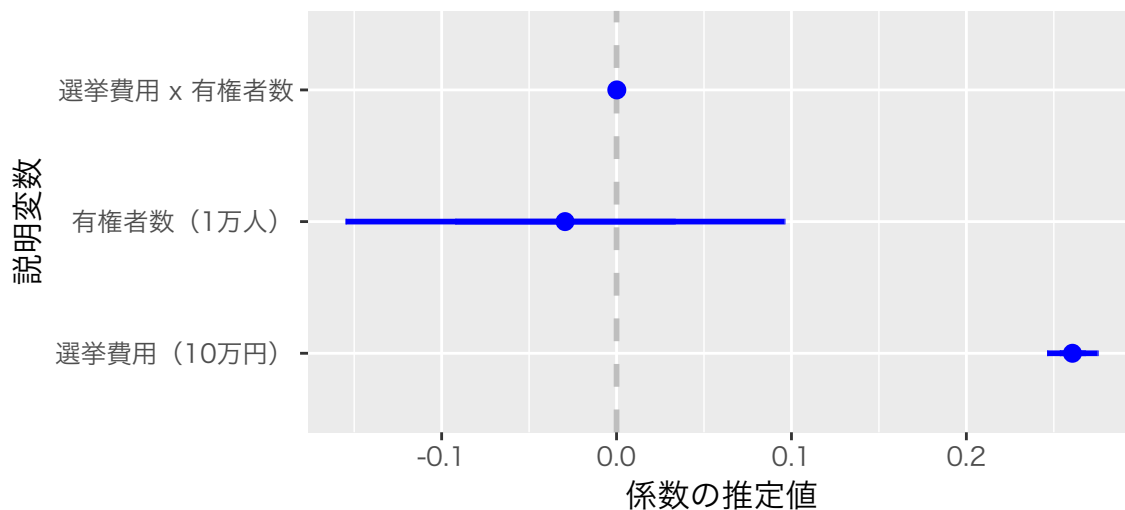
```
## lm(formula = voteshare ~ c_exp2 * c_eligible, data = HR_2005)
##               coef.est coef.se t value Pr(>|t|)
## (Intercept)    30.36     0.40   75.67   0.00
## c_exp2         0.26     0.01   36.17   0.00
## c_eligible     -0.03     0.06  -0.47   0.64
## c_exp2:c_eligible 0.00     0.00   0.13   0.90
## ---
## n = 985, k = 4
## residual sd = 12.58, R-Squared = 0.57
```

推定結果を図示する。

```
plt_fit_4 <- coefplot::coefplot(
  fit_4, intercept = FALSE, lwdOuter = 1,
  xlab = "係数の推定値", ylab = "説明変数",
  title = "回帰分析の結果：結果変数は得票率 (%)",
  newNames = c(c_exp2 = "選挙費用 (10 万円)",
               c_eligible = "有権者数 (1 万人)",
               `c_exp2:c_eligible` = "選挙費用 x 有権者数"))
print(plt_fit_4)
```



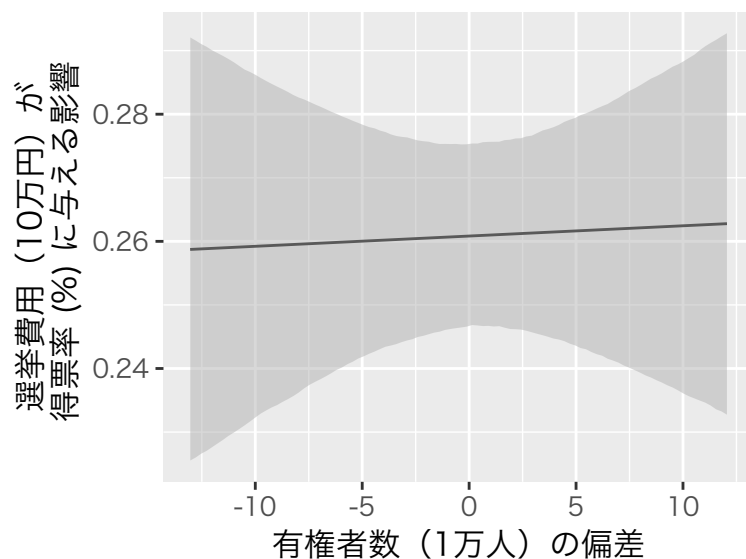
## 回帰分析の結果：結果変数は得票率 (%)



この図は、係数の点推定値と 95% 信頼区間を示している。この図を見る限り、選挙費用の効果は統計的に有意であるように見えるが、相互作用があるので、この図だけでは解釈が難しい。

そこで、選挙費用が得票率に与える影響が、有権者数によってどのように変化するか図示する。

```
plt_inter_4 <- interplot(fit_4, var1 = "c_exp2", var2 = "c_eligible") +
  labs(x = "有権者数 (1 万人) の偏差", y = "選挙費用 (10 万円) が\n 得票率 (%) に与える影響")
print(plt_inter_4)
```



この図は、選挙費用が得票率に与える影響が、有権者数によって変化するかどうか示している。この図を見る限り、選挙費用の影響は、有権者数にかかわらず常に正であることが読み取れる。また、有権者数は、「選挙費用が得票率に与える影響」に影響しているとは言えない。グレーで塗りつぶされた 95% 信頼区間を見ると、この区間内に傾きが正の直線も負の直線も引けてしまうことがわかる。つまり、有権者数の相互作用は統計的に有意ではない。

表 1 (最後のページ) のモデル 4 の列から、有権者数が平均値 (つまり、 $c\_eligible = 0$ ) の場合の効果量 (係数の推定値) を読むと、選挙費用の影響は 0.26 だということがわかる。つまり、選挙票を 100 万円増やすと、得票率が 2.6 ポイント増加することがわかる。この効果は、実質的にも有意な効果であると考えられる。

## (表を作る)

LaTeX を使って PDF を作るので、`texreg::texreg()` を使う。

```
texreg(list(fit_1, fit_2, fit_3, fit_4), single.row = FALSE,
  stars = numeric(0),
  custom.model.names = paste("モデル", 1:4),
  custom.coef.names = c("切片",
    "選挙費用 (10 万円)",
    "自民党ダミー",
    "選挙費用 x 自民党",
    "有権者数",
    "選挙費用 x 有権者数"),
  custom.note = "注：括弧内は標準誤差",
  include.rmse = FALSE,
  custom.gof.names = c("決定係数", "自由度調整済み決定係数", "観測数"),
  caption = "得票率を説明するモデルの推定結果",
  caption.above = TRUE, dcolumn = TRUE, use.packages = FALSE,
  label = "tbl:reg", float.pos = "h!")
```

表 1 得票率を説明するモデルの推定結果

	モデル 1	モデル 2	モデル 3	モデル 4
切片	30.35 (0.40)	26.23 (0.45)	27.40 (0.42)	30.36 (0.40)
選挙費用 (10 万円)	0.26 (0.01)	0.20 (0.01)	0.26 (0.01)	0.26 (0.01)
自民党ダミー		14.06 (0.93)	20.52 (0.95)	
選挙費用 x 自民党			-0.23 (0.02)	
有権者数				-0.03 (0.06)
選挙費用 x 有権者数				0.00 (0.00)
決定係数	0.57	0.65	0.71	0.57
自由度調整済み決定係数	0.57	0.65	0.71	0.57
観測数	985	985	985	985

注：括弧内は標準誤差