R第13回課題

2023-02-18 01:46:38

# 1.

jgss2000 <- read\_csv("Data/jgss2000\_v20210809.csv")

## Warning: One or more parsing issues, call `problems()` on your data frame for details,  
## e.g.:  
## dat <- vroom(...)  
## problems(dat)

## Rows: 2893 Columns: 376  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## dbl (372): iduse, block, pref, size, sexa, ageb, marc, dran, dobyear, xjob1w...  
## lgl (4): szpaywkx, szhilst, agemg2dv, agemg2wd  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

##男性ダミーの作成  
jgss2000 <- jgss2000 %>%   
 mutate(male\_dm = as.factor(2 - sexa))  
  
##変数の抜き出しと欠損値の除去  
jgss2000\_int <- jgss2000 %>%   
 select(income, educyr, male\_dm, ageb) %>%   
 drop\_na()  
  
##中心化  
jgss2000\_int <- jgss2000\_int %>%   
 mutate(ageb\_c = ageb - mean(ageb),  
 income\_c = income - mean(income),  
 educyr\_c = educyr - mean(educyr))  
  
##交互作用項を入れない分析  
model1 <- lm(income\_c ~ educyr\_c + male\_dm + ageb\_c,  
 data = jgss2000\_int)  
  
summary(model1)

##   
## Call:  
## lm(formula = income\_c ~ educyr\_c + male\_dm + ageb\_c, data = jgss2000\_int)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -583.19 -150.79 -31.92 91.37 2060.73   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -112.545 6.898 -16.32 < 2e-16 \*\*\*  
## educyr\_c 27.919 2.294 12.17 < 2e-16 \*\*\*  
## male\_dm1 253.898 10.408 24.40 < 2e-16 \*\*\*  
## ageb\_c -1.402 0.368 -3.81 0.000142 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 263.7 on 2640 degrees of freedom  
## Multiple R-squared: 0.274, Adjusted R-squared: 0.2731   
## F-statistic: 332.1 on 3 and 2640 DF, p-value: < 2.2e-16

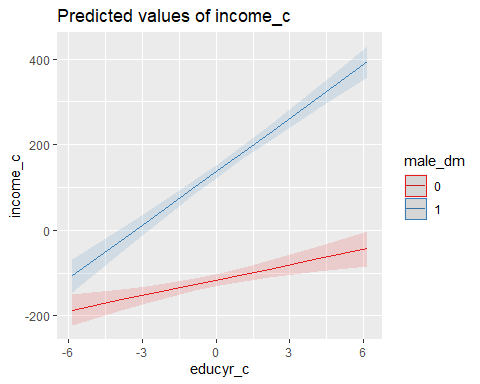
##交互作用項を入れた分析  
model2 <- lm(income\_c ~ male\_dm\*educyr\_c + ageb\_c,  
 data=jgss2000\_int)  
  
summary(model2)

##   
## Call:  
## lm(formula = income\_c ~ male\_dm \* educyr\_c + ageb\_c, data = jgss2000\_int)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -673.45 -138.55 -46.07 91.58 2014.99   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -116.7236 6.8425 -17.059 < 2e-16 \*\*\*  
## male\_dm1 253.4717 10.2922 24.628 < 2e-16 \*\*\*  
## educyr\_c 11.9621 3.0578 3.912 9.38e-05 \*\*\*  
## ageb\_c -1.7809 0.3671 -4.851 1.30e-06 \*\*\*  
## male\_dm1:educyr\_c 29.6904 3.8158 7.781 1.03e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 260.8 on 2639 degrees of freedom  
## Multiple R-squared: 0.2903, Adjusted R-squared: 0.2892   
## F-statistic: 269.8 on 4 and 2639 DF, p-value: < 2.2e-16

##結果  
library(modelsummary)  
cm <- c("(Intercept)" = "切片",  
 "male\_dm1" = "男性ダミー",  
 "educyr\_c" = "教育年数",  
 "ageb\_c" = "年齢",  
 "male\_dm1:educyr\_c" = "性別\*教育年数")  
modelsummary(list(model1,  
 model2),  
 fmt = 2,  
 stars = TRUE,  
 statistic = "conf.int",  
 coef\_map = cm,  
 gof\_map = c("nobs",  
 "r.squared",  
 "adj.r.squared"))

|  | (1) | (2) |
| --- | --- | --- |
| 切片 | -112.54\*\*\* | -116.72\*\*\* |
|  | [-126.07, -99.02] | [-130.14, -103.31] |
| 男性ダミー | 253.90\*\*\* | 253.47\*\*\* |
|  | [233.49, 274.31] | [233.29, 273.65] |
| 教育年数 | 27.92\*\*\* | 11.96\*\*\* |
|  | [23.42, 32.42] | [5.97, 17.96] |
| 年齢 | -1.40\*\*\* | -1.78\*\*\* |
|  | [-2.12, -0.68] | [-2.50, -1.06] |
| 性別\*教育年数 |  | 29.69\*\*\* |
|  |  | [22.21, 37.17] |
| Num.Obs. | 2644 | 2644 |
| R2 | 0.274 | 0.290 |
| R2 Adj. | 0.273 | 0.289 |
| + p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001 | | |

##交互作用効果の図示  
library(sjPlot)  
plot\_model(model2, type="pred", terms = c("educyr\_c","male\_dm"))



男女ともに教育年数は収入に影響を与えるが、女性よりも男性の方が、教育年数が収入に影響しやすいことがわかる。

## 感想・コメント・質問