

# HW1\_Wenquan\_Wu

## 1.1

普通话越好的人会不会越热衷于非正式政治参与。预期为普通话越好，越热衷于政治参与。

## 1.2

### 因变量

$y$ : D12c. 在这些活动或行动中，您是否担任过以下角色？

### 自变量

连续变量  $x_1$ : A8a 您个人去年全年的总收入是多少？

二元变量  $x_2$ : A10. 您目前的政治面貌是（党员，非党员）

分类变量  $x_3$ : A50. 您觉得自己说普通话的能力是什么水平？

## 1.3

```
#----load----
library(foreign)
library(DescTools)
library(dplyr)
library(lmtest)
library(BSDA)
options(scipen = 200, digits = 2)
data <- read.dta('./HW1/cgss2010_12.dta')
#----tidy_and_manipulate----
mydata <- data[,c('a49', 'a8a', 'a10', 'd12c')] %>%
  na.omit()

mydata <- mydata[mydata$a10 != '拒绝回答缺失值' &
  mydata$d12c != '拒绝回答' &
  mydata$d12c != '不知道' &
  mydata$a49 != '拒绝回答缺失值' &
  mydata$a49 != '不知道缺失值', ]
mydata <- mydata %>% mutate( cpcer = ifelse(a10 == '共产党员', TRUE, FALSE)) %>%
  mutate(protest = ifelse( d12c == '从未参与', FALSE, TRUE))

mydata <- mydata[, c('a8a', 'a49', 'cpcer', 'protest')]
names(mydata) <- c('income', 'nation_lan', 'cpcer', 'protest')
mydata$income <- as.numeric(mydata$income)
```

## 1.4

```
m1 <- glm(protest ~ income + cpcer + nation_lan,
  family = binomial(link = "logit"), data = mydata)
m0 <- glm(protest ~ 1,
  family = binomial(link = "logit"), data = mydata)

summary(m1)
summary(m0)
```

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.094   -0.853   -0.720    1.485    1.914

Coefficients:
              Estimate      Std. Error
(Intercept)  -0.199659689807  0.474978747678
income        0.000000000895  0.000000019328
cpcerTRUE     -0.438262949629  0.176464926801
nation_lan比较差 -0.499522649460  0.548361066067
nation_lan一般  -0.553700438598  0.491429229982
nation_lan比较好 -0.625061739427  0.486309391028
nation_lan很好  -1.018962151945  0.486768053636
              z value Pr(>|z|)
(Intercept)    -0.42    0.674
income          0.05    0.963
cpcerTRUE      -2.48    0.013 *
nation_lan比较差 -0.91    0.362
nation_lan一般  -1.13    0.260
nation_lan比较好 -1.29    0.199
nation_lan很好  -2.09    0.036 *
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1662.5 on 1430 degrees of freedom
Residual deviance: 1639.6 on 1424 degrees of freedom
AIC: 1654

Number of Fisher Scoring iterations: 4

```

截距 $\alpha$ 为-0.2,  $\beta_1$ 为0,  $\beta_2$ 为-0.44,  $\beta_3$ 为-0.5,  $\beta_4$ 为-0.55,  $\beta_5$ 为-0.62,  $\beta_6$ 为-1.01

## 1.5

```
anova(m1, test="Chisq")
```

```

              Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                                1430      1663
income          1      0.00      1429      1663    0.9712
cpcer           1      8.69      1428      1654    0.0032 **
nation_lan      4     14.23      1424      1640    0.0066 **
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

## 1.6

```
PseudoR2(m1)
```

```

McFadden
0.014

```

PseudoR2表示有自变量m1模型比没有自变量的m0模型可以解释 $y$ 更多的变化（多0.014）

## 1.7

```
z.test(m1, sigma.x=0.5, conf.level=0.95)
```

## 1.8

常数项：当income, pcer, nation\_lan都等于0时，你会参与游行抗议的概率为 $\text{logit}-1$   $(-0.19) = 0.45$ ，如果你不是党员，完全不会讲普通话，你会参与游行抗议的概率45%。

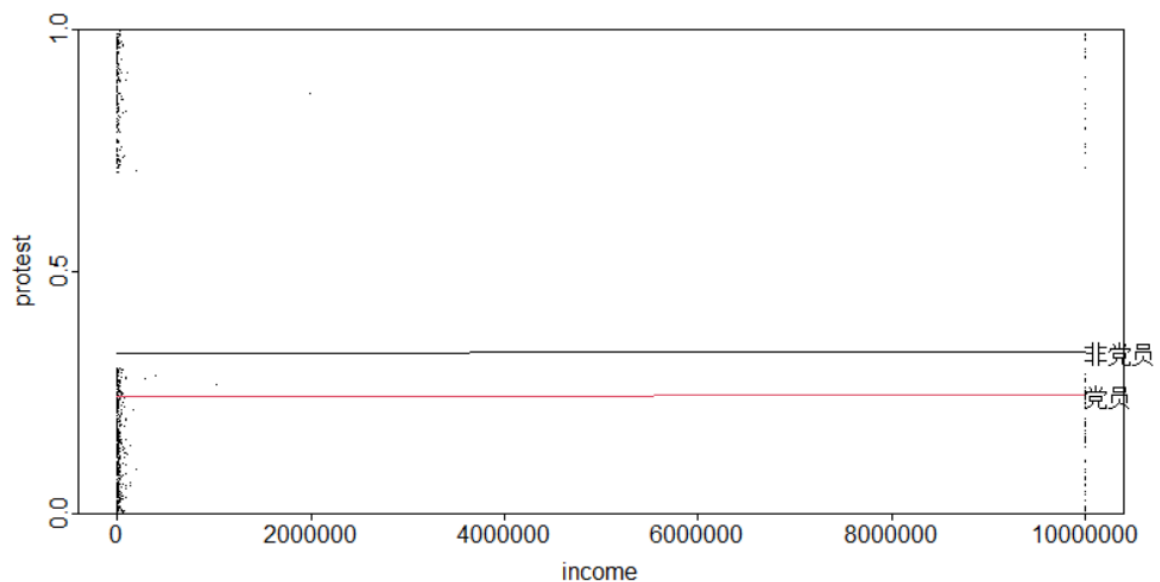
回归系数：

- 如果你是党员，你会参与游行抗议的概率就下降 $0.43/4 \approx 11\%$ 。
- 无论收入水平如何变化，你参与游行示威的水平受收入的影响很小。
- 如果你普通话属于比较差的水平，你会参与游行抗议的概率较完全不会普通话的人就会下降 $0.49/4 \approx 12\%$ 。
- 如果你普通话属于一般的水平，你会参与游行抗议的概率较完全不会普通话的人就会下降 $0.55/4 \approx 14\%$ 。
- 如果你普通话属于比较好的水平，你会参与游行抗议的概率较完全不会普通话的人就会下降 $0.62/4 \approx 15\%$ 。
- 如果你普通话属于很好的水平，你会参与游行抗议的概率较完全不会普通话的人就会下降 $1.01/4 \approx 25\%$ 。

## 1.9

```
invlogit<-function(x){
  1/(1+exp(-x))
}

par(mar=c(3,3,3,2), mgp=c(1.5,0.2,0), tcl=-0.2)
plot(x = income, y = protest, type = 'n', xlab = 'income', ylab = 'protest',
axes = FALSE,
      frame.plot = TRUE, xlim = c(0,9999999), yaxs = 'i')
points(x = income, y = jitter(as.numeric(protest), factor = 1.5), pch = '.')
axis(2, at=c(0,0.5,1))
axis(1)
curve(invlogit(coef(m1)[1] + coef(m1)[2] * x + coef(m1)[3] * TRUE + coef(m1)[4]
* 1), from = min(income), to = max(income), add = TRUE, col=2)
curve(invlogit(coef(m1)[1] + coef(m1)[2] * x + coef(m1)[3] * FALSE + coef(m1)[4]
* 1), from = min(income), to = max(income), add = TRUE, col=1)
text(x = max(income), y = invlogit( coef(m1)[1] + coef(m1)[2] * max(income) +
coef(m1)[3] * c(0,1) + coef(m1)[4] * 1), labels=c("非党员", "党员"),
      adj=0, xpd=NA)
```



## interaction term

### 2.4

```
m2 <- glm(protest ~ income + cpcer + nation_lan + cpcer:income,
           family = binomial(link = "logit"), data = mydata)
summary(m2)
```

#### Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.095	-0.855	-0.721	1.495	1.926

#### Coefficients:

	Estimate	Std. Error
(Intercept)	-0.19801199937	0.47506687498
income	-0.00000000289	0.00000002067
cpcerTRUE	-0.47027681693	0.18724155967
nation_lan比较差	-0.49490910751	0.54850318562
nation_lan一般	-0.55200569424	0.49151442608
nation_lan比较好	-0.62132865624	0.48643065708
nation_lan很好	-1.01687479573	0.48686031410
income:cpcerTRUE	0.00000003120	0.00000005763

	z value	Pr(> z )
(Intercept)	-0.42	0.677
income	-0.14	0.889
cpcerTRUE	-2.51	0.012 *
nation_lan比较差	-0.90	0.367
nation_lan一般	-1.12	0.261
nation_lan比较好	-1.28	0.201
nation_lan很好	-2.09	0.037 *
income:cpcerTRUE	0.54	0.588

---

#### Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 1662.5 on 1430 degrees of freedom
Residual deviance: 1639.3 on 1423 degrees of freedom
AIC: 1655

Number of Fisher Scoring iterations: 4
```

截距 $\alpha$ 为-0.2,  $\beta_1$ 为0,  $\beta_2$ 为-0.47,  $\beta_3$ 为-0.49,  $\beta_4$ 为-0.55,  $\beta_5$ 为-0.62,  $\beta_6$ 为-1.02,  $\beta_7$ 为0

## 2.5

```
anova(m2, test="Chisq")
```

```
              Df Deviance Resid. Df Resid. Dev
NULL                                1430      1663
income              1      0.00      1429      1663
cpcer               1      8.69      1428      1654
nation_lan          4     14.23      1424      1640
income:cpcer        1      0.28      1423      1639

              Pr(>Chi)
NULL
income              0.9712
cpcer               0.0032 **
nation_lan          0.0066 **
income:cpcer        0.5945
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 2.6

```
PseudoR2(m2)
```

```
McFadden
0.014
```

PseudoR2表示有自变量m2模型比没有自变量的m0模型可以解释y更多的变化（多0.014）

## 2.7

```
z.test(m2, sigma.x=0.5, conf.level=0.95)
```

## 2.8

常数项：当income, pcer, nation\_lan都等于0时，你会参与游行抗议的概率为 $\text{logit}^{-1}(-0.19) = 0.4$ ，如果你不是党员，完全不会讲普通话，你会参与游行抗议的概率0.4%。

回归系数：

- 如果你是党员，将收入水平定在均值1114225，你会参与游行抗议的概率就下降 $0.47/4 \approx 12\%$ 。
- 无论收入水平如何变化，将党员身份定在均值0.17，你参与游行示威的水平受收入的影响很小。
- 从党员身份的角度来看，增加个人收入，对于党员身份对于参与游行抗议这件事的概率影响很小；
- 从个人收入的角度来看，成为党员，对于个人收入对于参与游行抗议这件事的概率影响很小；
- 如果你普通话属于比较差的水平，你会参与游行抗议的概率较完全不会普通话的人就会下降 $0.49/4 \approx 12\%$ 。

- 如果你普通话属于一般的水平，你会参与游行抗议的概率较完全不会普通话的人就会下降 $0.55/4 \approx 14\%$ 。
- 如果你普通话属于比较好的水平，你会参与游行抗议的概率较完全不会普通话的人就会下降 $0.62/4 \approx 15\%$ 。
- 如果你普通话属于很好的水平，你会参与游行抗议的概率较完全不会普通话的人就会下降 $1.01/4 \approx 25\%$ 。

## 2.9

```
par(mar=c(3,3,3,2), mgp=c(1.5,0.2,0), tcl=-0.2)
plot(x = income, y = protest, type = 'n', xlab = 'income', ylab = 'protest',
axes = FALSE,
      frame.plot = TRUE, xlim = c(0,9999999), yaxs = 'i')
points(x = income, y = jitter(as.numeric(protest), factor = 1.5), pch = '.')
axis(2, at=c(0,0.5,1))
axis(1)
curve(invlogit(coef(m2)[1] + coef(m2)[2] * x + coef(m2)[3] * TRUE + coef(m2)[4]
* 1 + coef(m2)[8] * 1 * x), from = min(income), to = max(income), add = TRUE,
col=2)
curve(invlogit(coef(m2)[1] + coef(m2)[2] * x + coef(m2)[3] * FALSE + coef(m2)[4]
* 1), from = min(income), to = max(income), add = TRUE, col=1)
text(x = max(income), y = invlogit( coef(m1)[1] + coef(m1)[2] * max(income) +
coef(m1)[3] * c(0,1) + coef(m1)[4] * 1 +coef(m2)[8] * c(0,1) * max(income)),
labels=c("非党员", "党员"), adj=0, xpd=NA)
```

![image-20210308014213963]

(C:\Users\mi\AppData\Roaming\Typora\typora-user-images\image-20210308014213963.png)

