Pol Sci 630: Problem Set 7 - Dummy Variables and Interactions (II) - Solutions

Prepared by: Jan Vogler (jan.vogler@duke.edu)

Grading Due Date: Friday, October 21st, 1.40 PM (Beginning of Lab)

Insert your comments on the assignment that you are grading above the solution in bold and red text. For example write: "GRADER COMMENT: everything is correct!" Also briefly point out which, if any, problems were not solved correctly and what the mistake was. See below for more examples.

In order to make your text bold and red, you need to insert the following line at the beginning of the document:

```
\usepackage{color}
```

and the following lines above the solution of the specific task:

```
\textbf{\color{red} GRADER COMMENT: everything is correct!}
```

R Programming

Problem 1

```
library(foreign)
vote1 = read.dta("VOTE1.dta")
summary(vote1)

##     state               district          democA            voteA     
##  Length:173         Min.   : 1.000   Min.   :0.0000   Min.   :16.0  
##  Class :character   1st Qu.: 3.000   1st Qu.:0.0000   1st Qu.:36.0  
##  Mode  :character   Median : 6.000   Median :1.0000   Median :50.0  
##                     Mean   : 8.838   Mean   :0.5549   Mean   :50.5  
##                     3rd Qu.:11.000   3rd Qu.:1.0000   3rd Qu.:65.0  
##                     Max.   :42.000   Max.   :1.0000   Max.   :84.0  
##     expendA            expendB            prtystrA        lexpendA     
##  Min.   :   0.302   Min.   :   0.93   Min.   :22.00   Min.   :-1.197  
##  1st Qu.:  81.634   1st Qu.:  60.05   1st Qu.:44.00   1st Qu.: 4.402  
##  Median : 242.782   Median : 221.53   Median :50.00   Median : 5.492  
##  Mean   : 310.611   Mean   : 305.09   Mean   :49.76   Mean   : 5.026  
##  3rd Qu.: 457.410   3rd Qu.: 450.72   3rd Qu.:56.00   3rd Qu.: 6.126  
##  Max.   :1470.674   Max.   :1548.19   Max.   :71.00   Max.   : 7.293  
##     lexpendB            shareA       
##  Min.   :-0.07257   Min.   : 0.09464  
##  1st Qu.: 4.09524   1st Qu.:18.86800  
##  Median : 5.40056   Median :50.84990  
##  Mean   : 4.94437   Mean   :51.07654  
##  3rd Qu.: 6.11084   3rd Qu.:84.25510  
##  Max.   : 7.34484   Max.   :99.49500  

# Regular model

lm_vote = lm(voteA ~ expendA + expendB + prtystrA, data = vote1)
```

```r
summary(lm_vote)

##
## Call:
## lm(formula = voteA ~ expendA + expendB + prtystrA, data = vote1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -26.661  -8.385   0.362   8.536  30.814
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.267190   4.416784   7.532 2.87e-12 ***
## expendA      0.034924   0.003369  10.365  < 2e-16 ***
## expendB     -0.034924   0.003001 -11.636  < 2e-16 ***
## prtystrA     0.342514   0.087952   3.894 0.000142 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.12 on 169 degrees of freedom
## Multiple R-squared:  0.5687,Adjusted R-squared:  0.561
## F-statistic: 74.27 on 3 and 169 DF,  p-value: < 2.2e-16

# Regular model

lm_vote_fe = lm(voteA ~ expendA + expendB + prtystrA + factor(state) - 1, data = vote1)

summary(lm_vote_fe)

##
## Call:
## lm(formula = voteA ~ expendA + expendB + prtystrA + factor(state) -
##     1, data = vote1)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -28.701  -3.084   0.000   3.564  15.507 
## 
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)    
## expendA          0.006039   0.003290   1.836 0.068747 .  
## expendB         -0.006502   0.003100  -2.098 0.037933 *  
## prtystrA         0.358758   0.066177   5.421 2.90e-07 ***
## factor(state)AK 39.309023   8.357386   4.704 6.61e-06 ***
## factor(state)AL 51.365217   7.838766   6.553 1.30e-09 ***
## factor(state)AR 49.972076   6.379377   7.833 1.70e-12 ***
## factor(state)AZ 48.483292   6.488110   7.473 1.16e-11 ***
## factor(state)CA 43.048902   4.172163  10.318  < 2e-16 ***
## factor(state)CO 47.736154   4.974615   9.596  < 2e-16 ***
## factor(state)CT 46.210515   5.538809   8.343 1.09e-13 ***
## factor(state)DE 51.529573   7.864187   6.552 1.31e-09 ***
## factor(state)FL 44.776069   4.369228  10.248  < 2e-16 ***
## factor(state)GA 48.109371   4.031228  11.934  < 2e-16 ***
## factor(state)IA 44.698629   6.129685   7.292 2.98e-11 ***
## factor(state)ID 49.474737   7.700469   6.425 2.46e-09 ***
## factor(state)IL 43.231135   4.194895  10.306  < 2e-16 ***
## factor(state)IN 42.840843   4.509151   9.501  < 2e-16 ***
## factor(state)KS 51.489756   5.296654   9.721  < 2e-16 ***
## factor(state)KY 45.789701   5.071741   9.028 2.50e-15 ***
## factor(state)MA 50.339783   4.988827  10.091  < 2e-16 ***
## factor(state)MD 51.607950   8.165755   6.320 4.13e-09 ***
## factor(state)ME 45.491267   8.164868   5.572 1.46e-07 ***
## factor(state)MI 41.383011   3.937033  10.511  < 2e-16 ***
## factor(state)MN 18.089847   4.397459   4.114 6.98e-05 ***
## factor(state)MO 23.880602   4.643106   5.143 1.00e-06 ***
## factor(state)MT 25.107351   6.115619   4.105 7.20e-05 ***
## factor(state)NC 24.594941   4.602580   5.344 4.12e-07 ***
## factor(state)ND  9.475057   8.259547   1.147 0.253487    
```

```
## factor(state)NE 22.427537   6.245667   3.591 0.000471 ***
## factor(state)NJ 19.288478   5.119874   3.767 0.000252 ***
## factor(state)NM 22.318097   6.031893   3.700 0.000321 ***
## factor(state)NV 18.539845   8.443633   2.196 0.029943 *
## factor(state)NY 20.985113   3.970965   5.285 5.37e-07 ***
## factor(state)OH 12.289058   3.991892   3.079 0.002554 **
## factor(state)OK 23.895229   5.134888   4.654 8.14e-06 ***
## factor(state)OR 12.962933   7.969783   1.627 0.106340
## factor(state)PA 20.600347   4.203618   4.901 2.88e-06 ***
## factor(state)RI 20.859243   8.370642   2.492 0.014003 *
## factor(state)SC 24.502274   5.821299   4.209 4.84e-05 ***
## factor(state)SD 11.891797   8.260897   1.440 0.152481
## factor(state)TN  4.993908   8.493487   0.588 0.557605
## factor(state)TX 22.509159   4.373663   5.147 9.90e-07 ***
## factor(state)UT 27.349155   6.026371   4.538 1.31e-05 ***
## factor(state)VA 17.182274   6.008925   2.859 0.004968 **
## factor(state)WA 21.080793   4.831478   4.363 2.64e-05 ***
## factor(state)WI 20.782951   4.992736   4.163 5.79e-05 ***
## factor(state)WV 16.364273   6.003490   2.726 0.007328 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.33 on 126 degrees of freedom
## Multiple R-squared:  0.9862,Adjusted R-squared:  0.981
## F-statistic: 191.3 on 47 and 126 DF,  p-value: < 2.2e-16
```

a)

b) There are clear differences between the regular model and the model that uses fixed effects. While the direction of all coefficients stays the same, meaning that it is positive for incumbent expenditures, negative for challenger expenditures, and positive for party strength, we observe differences in both their absolute value and their statistical significance.

While the coefficients of all three variables are significant at all common levels in the regular model, in the fixed effects the coefficient of incumbent expenditures is significant only at $\alpha < 0.1$ and the coefficient of challenger expenditures is significant at $\alpha < 0.05$.

However, the coefficient of incumbent party strength remains significant at all common levels ($\alpha < 0.001$).

The introduction of fixed effects means that we control for the state in which the election takes place. This has two important consequences. The first one is that we compare elections within states to each other by introducing a state-specific average. The second is that we introduce a number of dummy variables to our model that each represent one state.

Problem 2

```r
lm_vote_int = lm(voteA ~ expendA + expendB + prtystrA + prtystrA * expendA,
    data = vote1)

summary(lm_vote_int)
```

```
##
## Call:
## lm(formula = voteA ~ expendA + expendB + prtystrA + prtystrA *
##     expendA, data = vote1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.5474  -7.4084  -0.6797   7.6570  28.6013
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       9.0708741  6.2529997   1.451    0.149
## expendA           0.1202457  0.0168882   7.120 3.00e-11 ***
## expendB          -0.0339101  0.0028049 -12.089  < 2e-16 ***
## prtystrA          0.8233044  0.1243655   6.620 4.63e-10 ***
## expendA:prtystrA -0.0016459  0.0003201  -5.142 7.50e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 10.37 on 168 degrees of freedom
## Multiple R-squared:  0.6273,Adjusted R-squared:  0.6184
## F-statistic: 70.69 on 4 and 168 DF,  p-value: < 2.2e-16
```

a)

b) Please note:

1. IVS = incumbent vote share
2. IPS = incumbent party strength
3. IPE = incumbent party expenditures
4. CPE = challenger party expenditures

When IPS is at a value of 0, for a 1-unit increase in IPE, we would expect a 9.071-unit increase in IVS. The base term of IPE is statistically significant at all common levels ($\alpha < 0.001$).

Generally, for a 1-unit increase in IPE, we would expect a $9.071 - 0.002 * IPS$ increase in IVS. The interaction term of IPE and IPS is statistically significant at all common levels ($\alpha < 0.001$).

For a 1-unit increase in CPE, we would expect a 0.034 decrease in IVS. This relationship is statistically significant at all common levels ($\alpha < 0.001$).

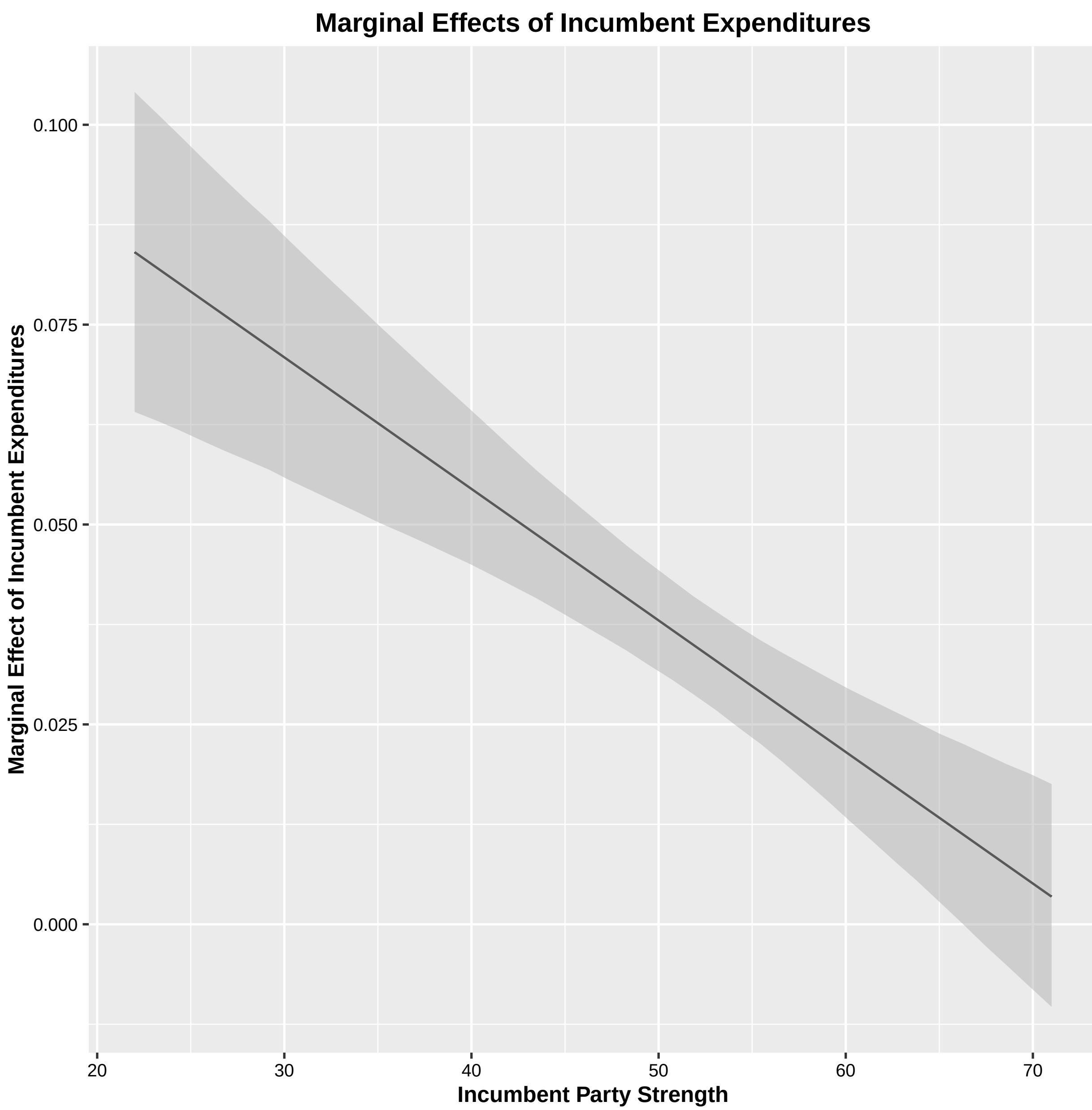
When IPS is at a value of 0, for a 1-unit increase in IPE, we would expect a 0.823 unit increase in IVS. The base term of IPE is statistically significant at all common levels ($\alpha < 0.001$).

Generally, for a 1-unit increase in IPS, we would expect a $0.823 - 0.002 * IPE$ increase in IVS. The interaction term of IPE and IPS is statistically significant at all common levels ($\alpha < 0.001$).

```r
library(interplot)
```

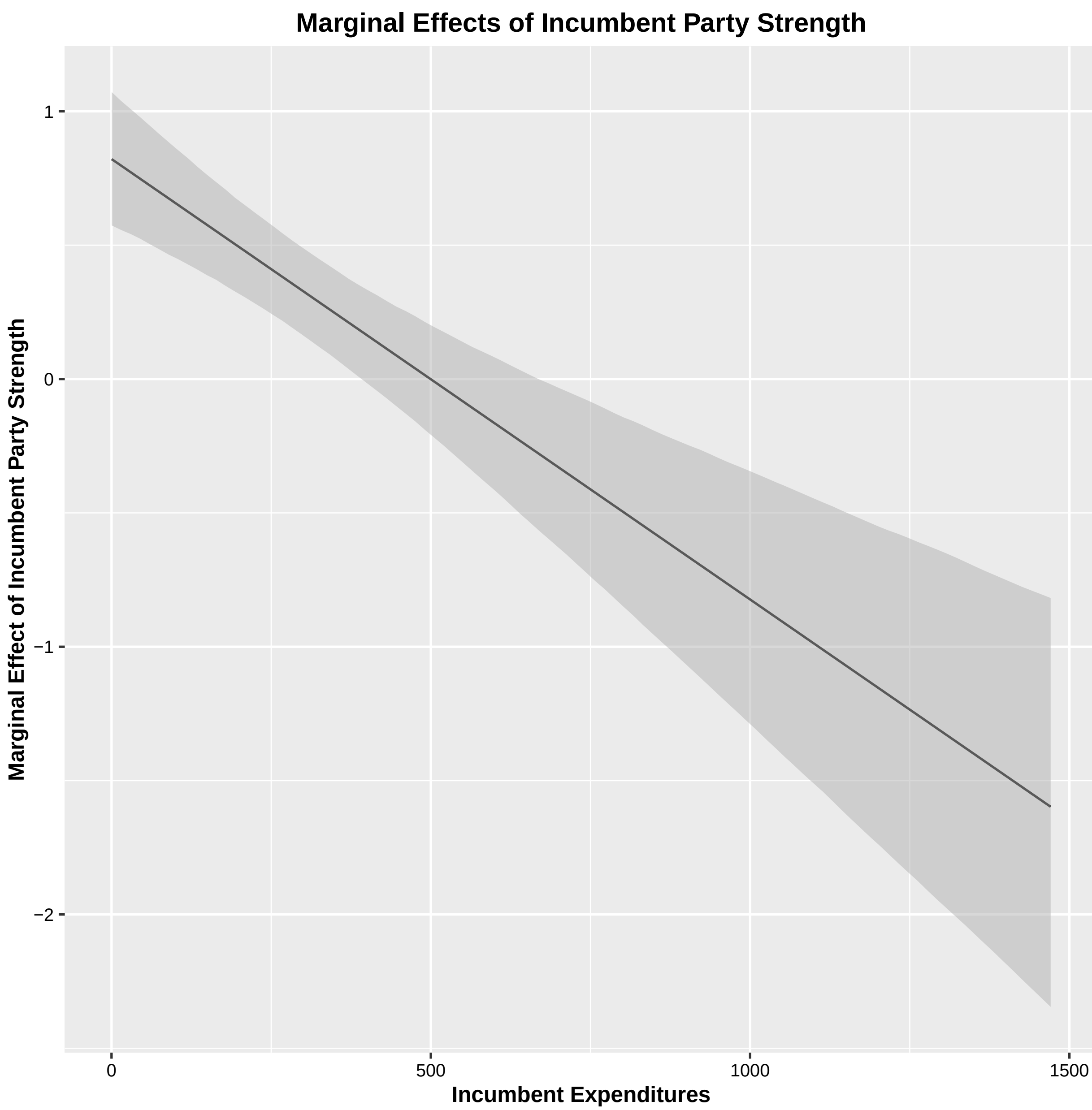
```
## Loading required package:  ggplot2
## Loading required package:  abind
## Loading required package:  arm
## Loading required package:  MASS
## Loading required package:  Matrix
## Loading required package:  lme4
##
## arm (Version 1.9-1, built:  2016-8-21)
## Working directory is C:/Users/Jan/OneDrive/Documents/GitHub/ps630_lab/ps630_f16/W7
```

```
interplot(m = lm_vote_int, var1 = "expendA", var2 = "prtystrA") + xlab("Incumbent Party
    ylab("Marginal Effect of Incumbent Expenditures") + ggtitle("Marginal Effects of Inc
    theme(plot.title = element_text(face = "bold", size = 12), axis.title = element_text
        face = "bold"), axis.text = element_text(size = 8, color = "Black"))
```



c)

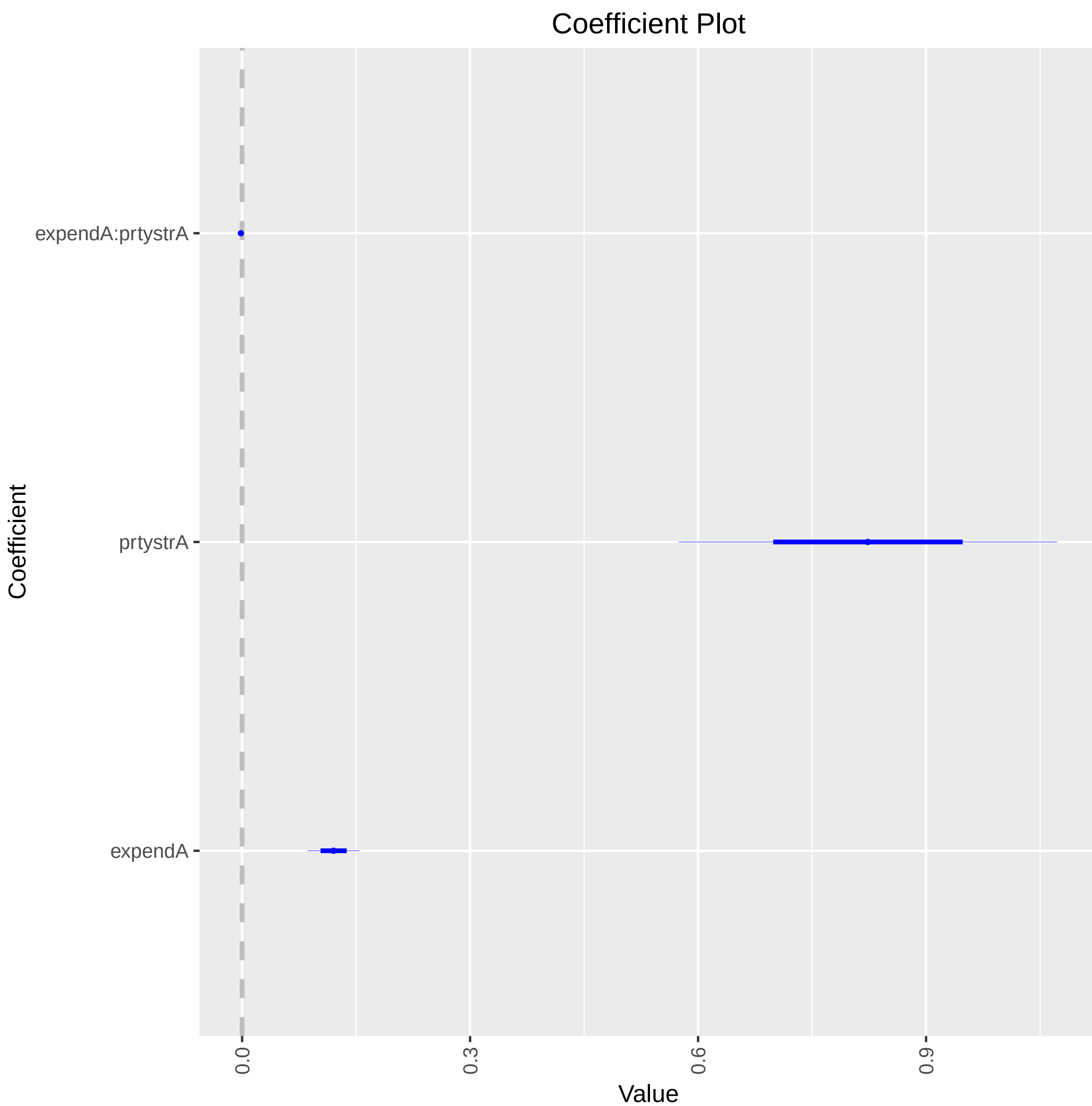
```
library(interplot)
interplot(m = lm_vote_int, var1 = "prtystrA", var2 = "expendA") + xlab("Incumbent Expend
    ylab("Marginal Effect of Incumbent Party Strength") + ggtitle("Marginal Effects of I
    theme(plot.title = element_text(face = "bold", size = 12), axis.title = element_text
        face = "bold"), axis.text = element_text(size = 8, color = "Black"))
```



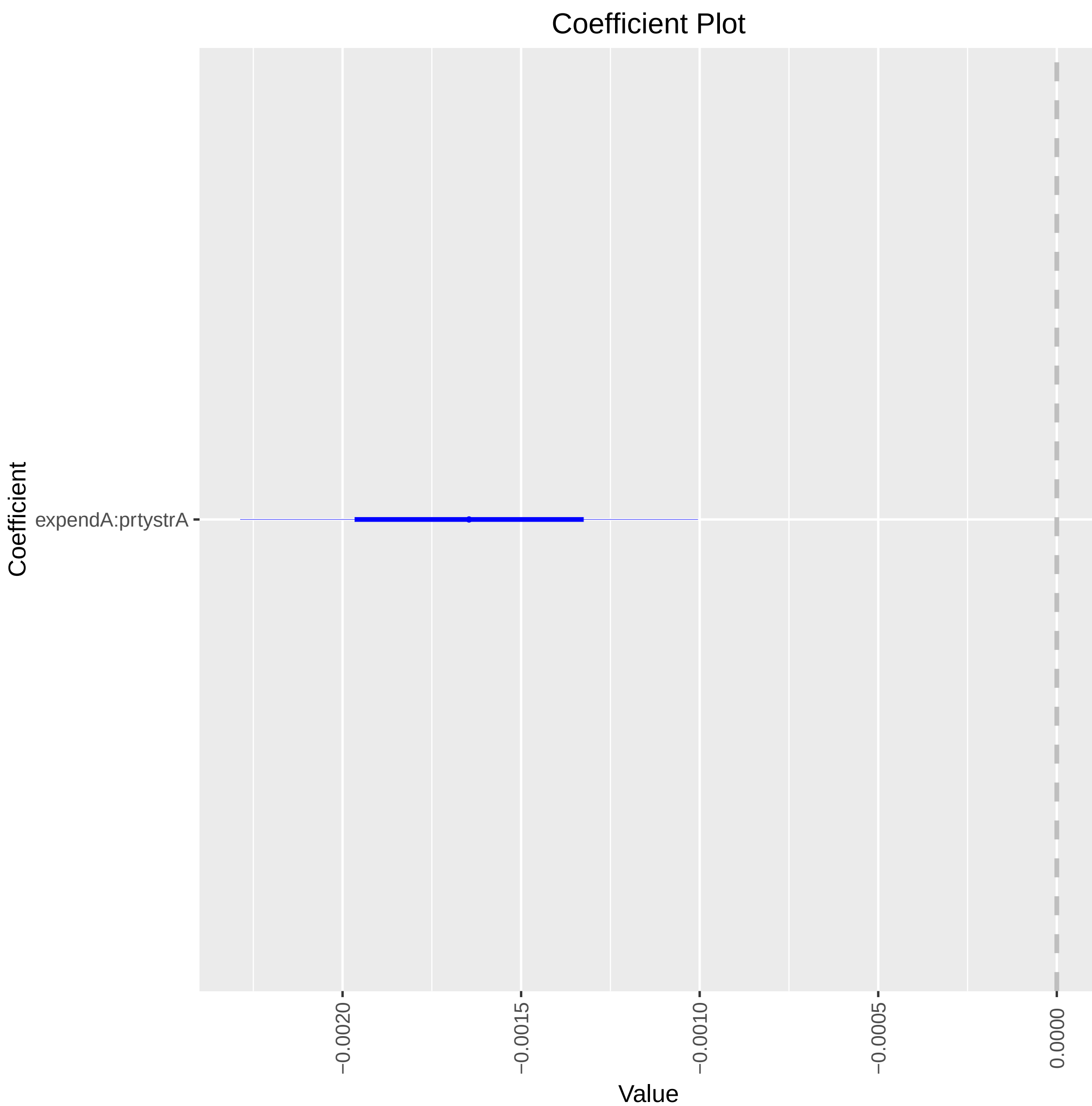

```
library(coefplot)

##
## Attaching package:  'coefplot'
## The following objects are masked from 'package:arm':
##
##     coefplot, coefplot.default

buildModelCI(lm_vote_int)
```

```
##                        Value      Coefficient    HighInner    LowInner
## expendA:prtystrA -0.00164590 expendA:prtystrA -0.001325801 -0.00196600
## prtystrA          0.82330445         prtystrA  0.947669935  0.69893896
## expendB          -0.03391007          expendB -0.031105130 -0.03671501
## expendA           0.12024567          expendA  0.137133916  0.10335743
## (Intercept)       9.07087413      (Intercept) 15.323873796  2.81787446
##                    HighOuter     LowOuter       Model
## expendA:prtystrA -0.001005701 -0.002286099 lm_vote_int
## prtystrA          1.072035425  0.574573465 lm_vote_int
## expendB          -0.028300190 -0.039519951 lm_vote_int
## expendA           0.154022157  0.086469192 lm_vote_int
## (Intercept)      21.576873462 -3.435125201 lm_vote_int
```

```
coefplot(lm_vote_int, coefficients = c("expendA", "prtystrA", "expendA:prtystrA"),
    point = T) + theme(axis.text.x = element_text(angle = 90))
```



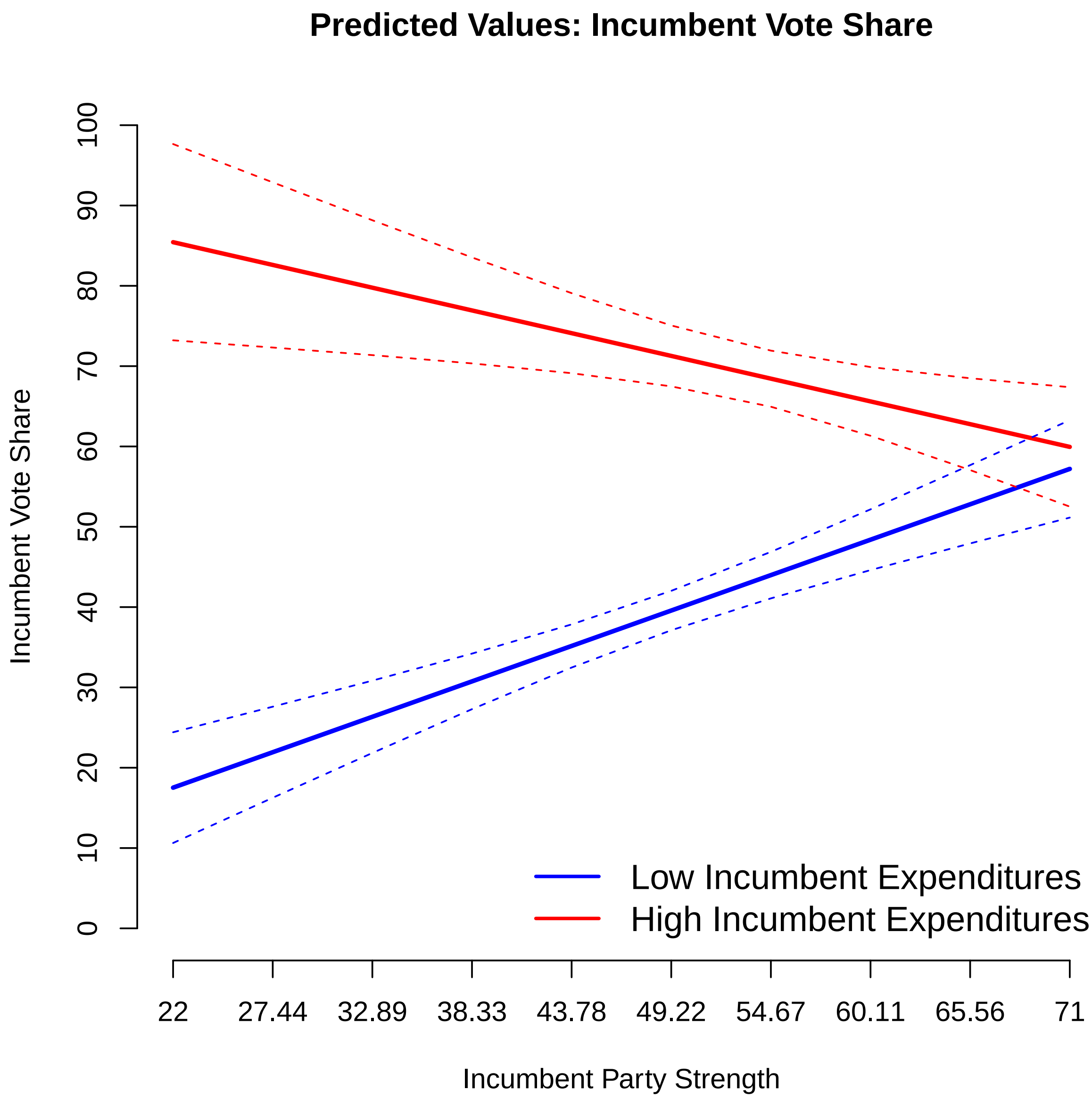

```
coefplot(lm_vote_int, coefficients = c("expendA:prtystrA"), point = T) + theme(axis.text
```




```r
quantile(vote1$expendA, probs = c(0.05, 0.95), na.rm = TRUE)

##       5%      95%
##   8.1354 816.2566

nd1 = data.frame(prtystrA = seq(min(vote1$prtystrA), max(vote1$prtystrA), length.out = 1
    expendA = rep(8.1354, 10), expendB = rep(mean(vote1$expendB), 10))
nd2 = data.frame(prtystrA = seq(min(vote1$prtystrA), max(vote1$prtystrA), length.out = 1
```

```
    expendA = rep(816.2566, 10), expendB = rep(mean(vote1$expendB), 10))

pred.p1 = predict(lm_vote_int, type = "response", se.fit = TRUE, newdata = nd1)
pred.p2 = predict(lm_vote_int, type = "response", se.fit = TRUE, newdata = nd2)

pred.table1 = cbind(pred.p1$fit, pred.p1$se.fit)
pred.table2 = cbind(pred.p2$fit, pred.p2$se.fit)

max(pred.table1)

## [1] 57.20747

max(pred.table2)

## [1] 85.43283

min(pred.table1)

## [1] 1.227114

min(pred.table2)

## [1] 1.746126

plot(pred.p1$fit, type = "l", ylim = c(0, 100), main = "Predicted Values: Incumbent Vote
    xlab = "Incumbent Party Strength", ylab = "Incumbent Vote Share", axes = FALSE,
    col = "blue", lwd = 2.5)
axis(1, at = seq(1, 10), labels = round(seq(min(vote1$prtystrA), max(vote1$prtystrA),
    length.out = 10), digits = 2))
axis(2, at = seq(0, 100, by = 10), labels = seq(0, 100, by = 10))

# Add lines

lines(pred.p1$fit, col = "blue", lwd = 2.5)
lines(pred.p2$fit, col = "red", lwd = 2.5)

# Add legend
```

```r
legend("bottomright", c("Low Incumbent Expenditures", "High Incumbent Expenditures"),
    lty = 1, lwd = 2, col = c("blue", "red"), bty = "n", cex = 1.25)

# Add confidence intervals

fit1 = pred.p1$fit
low1 = pred.p1$fit - 2 * pred.p1$se.fit
high1 = pred.p1$fit + 2 * pred.p1$se.fit
cis1 = cbind(fit1, low1, high1)

fit2 = pred.p2$fit
low2 = pred.p2$fit - 2 * pred.p2$se.fit
high2 = pred.p2$fit + 2 * pred.p2$se.fit
cis2 = cbind(fit2, low2, high2)

matlines(cis1[, c(2, 3)], lty = 2, col = "blue")
matlines(cis2[, c(2, 3)], lty = 2, col = "red")
```



d)

1. A lower level of incumbent party strength is associated with a more positive effect of incumbent party expenditures. This is so because, due to the negative coefficient of the interaction term, for decreases in incumbent party strength, we will see an increase in the marginal effect of incumbent party expenditures.

2. A higher level of incumbent party expenditures is associated with a more negative

effect of incumbent party strength. This is so because, due to the negative coefficient of the interaction term, for increases in incumbent party expenditures, we will see a decrease in the marginal effect of incumbent party strength.

Interactions: Math and Interpretation

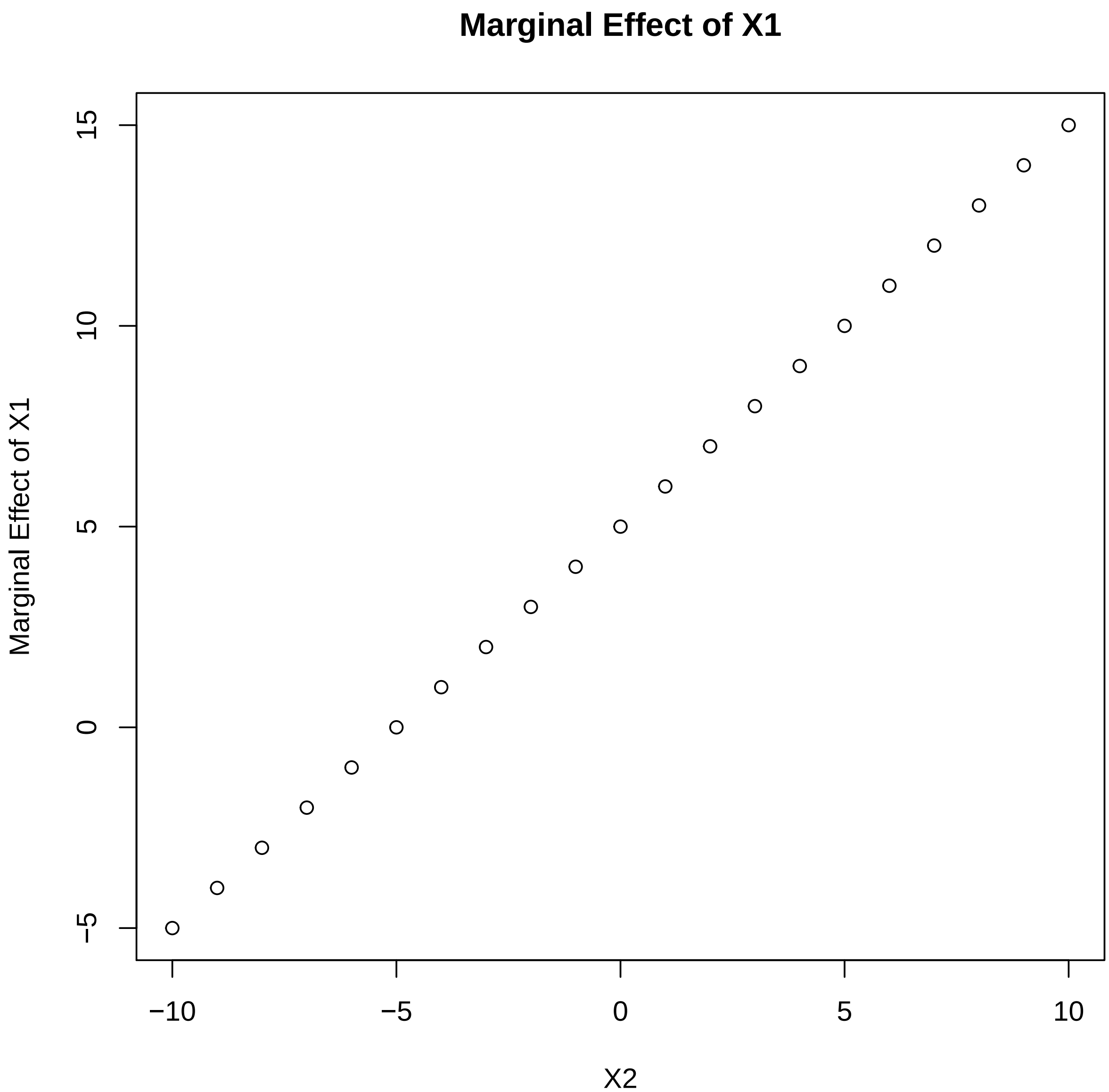
Problem 3

a)

$$\frac{\partial Y}{\partial X_1} = 5 + X_2$$

$$\frac{\partial Y}{\partial X_2} = 2 + X_1$$

```
x2 = seq(-10, 10)

marginal_effect_x1 = rep(5, 21) + x2

plot(x2, marginal_effect_x1, main = "Marginal Effect of X1", xlab = "X2", ylab = "Margin
```



b)

c) When X_2 is at a value of 0, for a 1-unit increase in X_1, we would expect a 5 unit increase in Y.

Generally, for a 1-unit increase in X_1, we would expect a $5 + X_2$ increase in Y.

d) If an interaction exists in reality and we omit the interaction term from our model, we will introduce a bias to our coefficients. The reason for this is that situations in which a high value in Y is caused by jointly high values in X_1 and X_2 will not be correctly captured by the model. Instead, high values in Y will incorrectly be attributed to either X_1 or X_2, but not to their interaction.