Answers to Problem Set 2

Nicholas Gonzalez

Worked on with Tanner Bentley.

Problem #1

```
library(haven)

zambia <- read_dta("zambia.dta")
```

Problem 1, Part A

```
voteMP_treatment <- lm(zambia$voteMP ~ zambia$treatment)

print(voteMP_treatment)
```

```

Call:
lm(formula = zambia$voteMP ~ zambia$treatment)

Coefficients:
     (Intercept)  zambia$treatment
         0.34211           0.03627
```

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
library(modelsummary)
```

```
`modelsummary` 2.0.0 now uses `tinytable` as its default table-drawing
  backend. Learn more at: https://vincentarelbundock.github.io/tinytable/

Revert to `kableExtra` for one session:

  options(modelsummary_factory_default = 'kableExtra')
  options(modelsummary_factory_latex = 'kableExtra')
  options(modelsummary_factory_html = 'kableExtra')

Silence this message forever:

  config_modelsummary(startup_message = FALSE)
```

```
modelsummary(voteMP_treatment)
```

Interpreting this model, or the effect, we see that someone being informed that their chief voted for the MP candidate, has a positive effect, and a magnitude of .036. For the intercept, the sign has a positive effective, with a slope or magnitude of change of 0.342. This would be 34.2%. Our treatment, which is being told the cheif's preferences.

Problem 1, Part B, i

```
interaction_model_1 <- lm(voteMP ~ treatment * chiefMPimportant, data = zambia)

modelsummary(interaction_model_1)
```

	(1)
(Intercept)	0.342
	(0.035)
zambia$treatment	0.036
	(0.050)
Num.Obs.	375
R2	0.001
R2 Adj.	−0.001
AIC	519.2
BIC	531.0
Log.Lik.	−256.596
RMSE	0.48

	(1)
(Intercept)	0.306
	(0.045)
treatment	−0.046
	(0.065)
chiefMPimportant	0.027
	(0.071)
treatment × chiefMPimportant	0.231
	(0.099)
Num.Obs.	368
R2	0.041
R2 Adj.	0.033
AIC	493.0
BIC	512.5
Log.Lik.	−241.488
RMSE	0.47

	(1)
(Intercept)	0.376
	(0.041)
treatment	0.021
	(0.057)
chiefchangetreat	−0.060
	(0.089)
treatment × chiefchangetreat	−0.014
	(0.128)
Num.Obs.	359
R2	0.004
R2 Adj.	−0.005
AIC	506.0
BIC	525.4
Log.Lik.	−248.008
RMSE	0.48

```
interaction_model_2 <- lm(voteMP ~ treatment * chiefchangetreat, data = zambia)

# View the summary of the model
modelsummary(interaction_model_2)
```

```
modelsummary(list("Model 1" = interaction_model_1, "Model 2" = interaction_model_2))
```

Model 1:

$$\text{voteMP}_i = \beta_0 + \beta_1 \cdot \text{treatment}_i + \beta_2 \cdot \text{chiefMPimportant}_i + \beta_3 \cdot (\text{treatment}_i \times \text{chiefMPimportant}_i) + u_i$$

Model 2:

$$\text{voteMP}_i = \beta_0 + \beta_1 \cdot \text{treatment}_i + \beta_2 \cdot \text{chiefchangetreat}_i + \beta_3 \cdot (\text{treatment}_i \times \text{chiefchangetreat}_i) + u_i$$

	Model 1	Model 2
(Intercept)	0.306	0.376
	(0.045)	(0.041)
treatment	−0.046	0.021
	(0.065)	(0.057)
chiefMPimportant	0.027	
	(0.071)	
treatment × chiefMPimportant	0.231	
	(0.099)	
chiefchangetreat		−0.060
		(0.089)
treatment × chiefchangetreat		−0.014
		(0.128)
Num.Obs.	368	359
R2	0.041	0.004
R2 Adj.	0.033	−0.005
AIC	493.0	506.0
BIC	512.5	525.4
Log.Lik.	−241.488	−248.008
RMSE	0.47	0.48

Problem 1, Part B, ii

Interpreting both models:

Model 1: Interaction Between Treatment and Chief-MP Importance

The intercept, or baseline for the respondents in the control group who did not consider both the chief, or MP important is .306. The treatment for the latter group has a negative sign, so it is decreasing, and a magnitude of .046 (so, -0.046). The treatment for this however is larger for the respondents who consider the MP and chief important. This tells us that the treatment of the experiment is more effective for this group. This highlights the perceived important of both the chief, as well as the MP.

Model 2: Interaction Between Treatment and Fear of Retaliation

For the control group that does not fear relation the sign is positive, and the magnitude is 0.376. The treatment effect for this specific group is .021, so also positive. The respondents who fear retaliation of some kind have an outcome that -0.060 lower, but the interaction between the treatment and fear of relation is smaller (-0.014). This tells us that fear of retaliation does not greatly affect the treatment, and the latter does not greatly vary across groups.

Problem 1, Part B, iii

```
pred_1 <- predict(interaction_model_1, newdata = zambia)

zambia$pred_1 <- pred_1
```

```
library(tidyverse)

# For interaction_model_1 (with `chiefMPimportant`)
table_1 <- zambia %>%
  group_by(treatment, chiefMPimportant) %>%
  summarise(mean_pred_1 = mean(pred_1), 3) %>%
  spread(key = chiefMPimportant, value = mean_pred_1)
```

```
`summarise()` has grouped output by 'treatment'. You can override using the
`.groups` argument.
```

```
# View the tables
print(table_1)
```

```
# A tibble: 3 x 5
# Groups:   treatment [3]
  treatment   `3`    `0`    `1` `<NA>`
      <dbl> <dbl>  <dbl>  <dbl>  <dbl>
1         0     3  0.306  0.333     NA
2         1     3  0.260  0.518     NA
3        NA     3 NA     NA         NA
```

```
library(tidyverse)
library(knitr)

table_1 <- zambia %>%
  group_by(treatment, chiefMPimportant) %>%
  summarise(mean_pred_1 = round(mean(pred_1), 3)) %>%
  spread(key = chiefMPimportant, value = mean_pred_1) %>%
  mutate(treatment = if_else(treatment == 0, "Control", "Treatment")) %>%
  rename(
    "Both  Important" = "0",
    "Both Not Important" = "1"
  )
```

```
`summarise()` has grouped output by 'treatment'. You can override using the
`.groups` argument.
```

```
kable(table_1,
      caption = "Mean Predictions by Treatment and Chief MP Importance",
      align = c("l", "c", "c"))
```

Table 1: Mean Predictions by Treatment and Chief MP Importance

treatment	Both Important	Both Not Important	
Control	0.306	0.333	NA
Treatment	0.260	0.518	NA
NA	NA	NA	NA

Problem 1, Part B, iv

```r
library(ggeffects)
library(ggplot2)
```

```r
interaction_model_article <- lm(voteMP ~ treatment * chiefMPimportant , data = zambia)
```

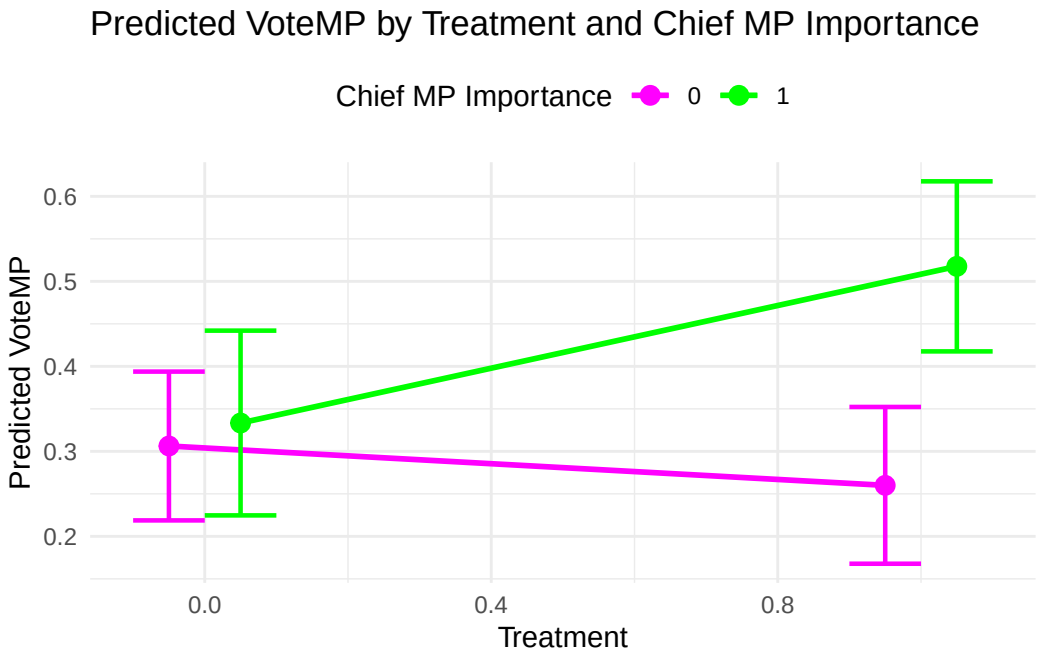
```r
interaction_model_article <- ggeffects::ggpredict(interaction_model_article, terms = c("treat

head(interaction_model_article)
```

```
# Predicted values of voteMP

chiefMPimportant: 0

treatment | Predicted |     95% CI
----------------------------------
        0 |      0.31 | 0.22, 0.39
        1 |      0.26 | 0.17, 0.35

chiefMPimportant: 1

treatment | Predicted |     95% CI
----------------------------------
        0 |      0.33 | 0.22, 0.44
        1 |      0.52 | 0.42, 0.62
```

```r
library(ggplot2)

ggplot(interaction_model_article, aes(x = x, y = predicted, color = group)) +
  geom_point(size = 3, position = position_dodge(width = 0.2)) +
  geom_line(aes(group = group), size = 1, position = position_dodge(width = 0.2)) +
  geom_errorbar(aes(ymin = conf.low, ymax = conf.high),
                width = 0.2,  # Width of the error bar ends
                position = position_dodge(width = 0.2),
                size = 0.8) +  # Thickness of the error bars
  labs(
    title = "Predicted VoteMP by Treatment and Chief MP Importance",
    x = "Treatment",
    y = "Predicted VoteMP",
    color = "Chief MP Importance"
  ) +
```

```
  scale_color_manual(values = c("magenta", "green")) +
  theme_minimal() +
  theme(legend.position = "top")
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```



Problem 2, Part A

```
load("~/Documents/PS405-Linear-Models/Problem_Set_2/schools.RData")
```

```
model_schools <- lm(math ~ size.small + size.medium + size.large, data = schools)

model_schools
```

```

Call:
lm(formula = math ~ size.small + size.medium + size.large, data = schools)

Coefficients:
```

	(1)
(Intercept)	19.747
	(0.238)
size.small	5.094
	(0.362)
size.medium	3.230
	(0.275)
Num.Obs.	300
R2	0.424
R2 Adj.	0.420
AIC	1238.3
BIC	1253.1
Log.Lik.	−615.132
RMSE	1.88

```
(Intercept)   size.small  size.medium   size.large
     19.747        5.094        3.230           NA
```

```
modelsummary(model_schools)
```

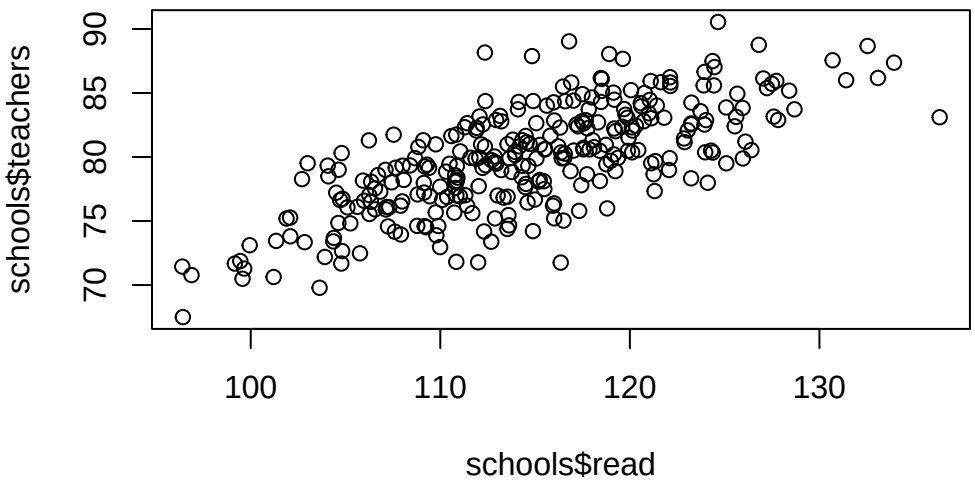
The results automatically picked the reference category as "size.large" so that is what the intercept is. R/OLS automatically picks a reference category. If we look at our model it is why it says NA for size.large. The other categories, size.small and size.medium, are being compared to size.large since that is the reference category. Both other sizes, medium and small have positive slope coefficients, with magnitudes of 5.094 and 3.230 respectively.

The original model shown uses all three dummy, or categorical variables, those being small, medium and large. This means we do not have a reference group(technically, but R fixes this for us). To fix the model's notation, we would drop one categorical variable, to have it serve as a reference group. For this model, I picked *small.* This violates the multi-collinearity assumption/rule.

$$\hat{\text{math}}_i = \beta_0 + \beta_1 \cdot \text{size.medium}_i + \beta_2 \cdot \text{size.large}_i + u_i$$

Problem 2, Part B

```
plot(schools$read, schools$teachers)
```



The data looks like it could be non-linear, but the result is inconclusive from looking at this scatterplot. There is some evidence of a quadratic shape.

```
linear_schools <- lm(read ~ teachers + size.medium + size.large, data = schools)

modelsummary(linear_schools)
```

```
poly_schools <- lm(read ~ poly(teachers, 2) + size.medium + size.large, data = schools)


modelsummary(poly_schools)
```

```
modelsummary(list("Linear" = linear_schools, "Poly" = poly_schools))
```

The polynomial model created accounts for the non-linear features in the relationship between the teachers, and reading scores. However, the slope coefficients for the medium and large schools, in both the polynomial and linear model are similar, which may suggest that making this a polynomial model does not change the estimated effects significantly. This can also be inferred via the R^2, which are both very similar. A benefit of the polynomial model however is that we can see the impact the number of teachers has. The results of the model suggest that there is a nonlinear relationship between amount of teachers and outcome as the model

	(1)
(Intercept)	16.103
	(5.710)
teachers	1.253
	(0.071)
size.medium	−1.575
	(0.814)
size.large	−2.532
	(0.967)
Num.Obs.	300
R2	0.526
R2 Adj.	0.521
AIC	1827.6
BIC	1846.1
Log.Lik.	−908.803
RMSE	5.00

shows a higher number of teachers has a larger and growing effect on the outcome. This is something sort of inferred previously in the scatterplot.

“‘

	(1)
(Intercept)	116.338
	(0.729)
poly(teachers, 2)1	89.596
	(5.056)
poly(teachers, 2)2	−6.224
	(5.061)
size.medium	−1.660
	(0.817)
size.large	−2.653
	(0.971)
Num.Obs.	300
R2	0.529
R2 Adj.	0.522
AIC	1828.1
BIC	1850.3
Log.Lik.	−908.036
RMSE	4.99

	Linear	Poly
(Intercept)	16.103	116.338
	(5.710)	(0.729)
teachers	1.253	
	(0.071)	
size.medium	−1.575	−1.660
	(0.814)	(0.817)
size.large	−2.532	−2.653
	(0.967)	(0.971)
poly(teachers, 2)1		89.596
		(5.056)
poly(teachers, 2)2		−6.224
		(5.061)
Num.Obs.	300	300
R2	0.526	0.529
R2 Adj.	0.521	0.522
AIC	1827.6	1828.1
BIC	1846.1	1850.3
Log.Lik.	−908.803	−908.036
RMSE	5.00	4.99