

Diversity & Innovation in Academia

Gloria Grace

September 15, 2022

Data

```
pi <- read_csv('https://sldr.netlify.app/data/phd_innovation.csv')
```

```
## Rows: 4195 Columns: 8
```

```
## -- Column specification -----
```

```
## Delimiter: ","
```

```
## chr (3): field, race_ethnicity, sex
## dbl (5): year, prop_same_race, prop_same_sex, novelty, uptake
##
## 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
```

Fit Linear Regression

```
mlr <- lm(uptake ~ prop_same_sex + sex, data = pi )
```

Model Summary

```
summary(mlr)
```

```
##
## Call:
```

```
## lm(formula = uptake ~ prop_same_sex + sex, data = pi)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4596 -0.4377 -0.1111  0.3326  4.0992
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.81192    0.03622   22.419  <2e-16 ***
## prop_same_sex    0.95071    0.06638   14.322  <2e-16 ***
## sexMale         0.03620    0.02139    1.692   0.0906 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6427 on 4192 degrees of freedom
## Multiple R-squared:  0.05768,    Adjusted R-squared:  0.05723
## F-statistic: 128.3 on 2 and 4192 DF,  p-value: < 2.2e-16
```

Equation:

$$\text{uptake} = 0.81 + 0.95 \text{ prop_same_sex} + 0.04 \text{ Imale} + \epsilon,$$

$\epsilon \sim N(0, 0.64)$, Imale is 1 if pi is male, 0 otherwise.

Making a Prediction Plot

new prediction data

```
fake_data <- expand.grid(prop_same_sex = seq(from = 0.2,  
                                             to = 0.9,  
                                             by = 0.01),  
                        sex = c("Male"))
```

SE Predictions

```
preds <- predict(mlr,  
                 newdata = fake_data,  
                 se.fit = TRUE)
```

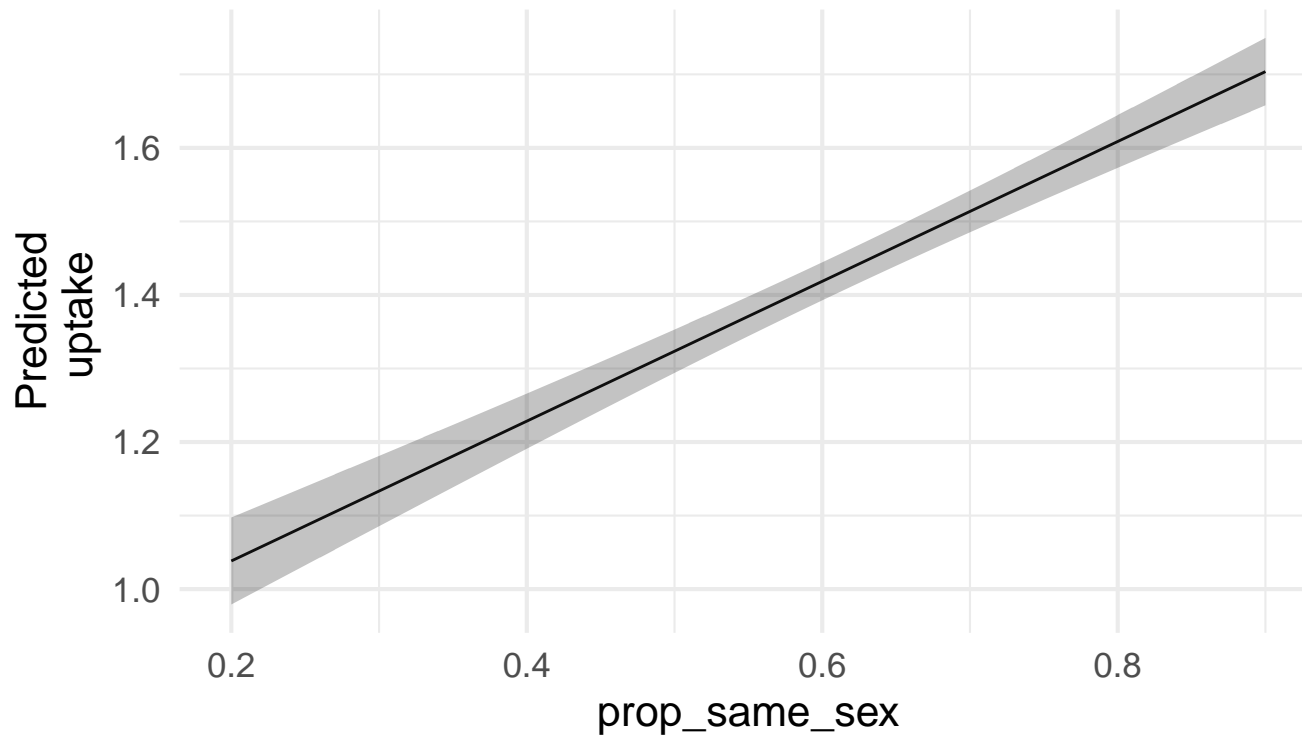
```
fake_data <- fake_data %>%  
  mutate(pred = preds$fit,  
         pred.se = preds$se.fit)
```

Convert from SE to CI

```
fake_data <- fake_data %>%  
  mutate(CI_lower = pred - 1.96*pred.se,  
         CI_upper = pred + 1.96*pred.se)
```

Prediction Plot

```
gf_line(pred ~ prop_same_sex,  
        data = fake_data) %>%  
  gf_labs(y = 'Predicted\n uptake') %>%  
  gf_ribbon(CI_lower + CI_upper ~ prop_same_sex)
```



For my fitted model, I chose uptake as my response because I want to investigate other factors with it and both sex and prop_same_sex as my predictors. I chose my model by looking at the data set and understanding what each variable represents.

From the prediction plot, I can take a conclusion that through this prediction, a measure of how enthusiastically innovative ideas are taken up or uptake are taken higher as proportion of peers who are the same sex also goes higher.