# Problem Set 3

Priyanshu Dey 703

2026-02-18

## Problem 5: Utility of Non-Linear Regression Over Linear Regression

### (a)

Considering the refractive index (RI) of "Vehicle Window glass" as the variable of interest and assuming linearity of regression, run multiple linear regression of RI on different metallic oxides.

From the p value, report which metallic oxide best explains the refractive index.

```
data("fgl")

df_veh = subset(fgl, type == "Veh")

model_a = lm(RI ~ Na + Mg + Al + Si + K + Ca + Ba + Fe, data = df_veh)

summary(model_a)

##
## Call:
## lm(formula = RI ~ Na + Mg + Al + Si + K + Ca + Ba + Fe, data = df_veh)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29194 -0.08582  0.00072  0.10740  0.33524
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 131.4641    47.2669   2.781  0.02388 *
## Na           -0.4333     0.3509  -1.235  0.25190
## Mg           -0.2866     1.0075  -0.285  0.78325
## Al           -0.8909     0.5550  -1.605  0.14713
## Si           -1.8824     0.4993  -3.770  0.00547 **
## K            -2.4232     0.9725  -2.492  0.03743 *
## Ca            1.5326     0.5818   2.634  0.02998 *
## Ba            0.3517     2.6904   0.131  0.89922
## Fe            3.8931     0.9581   4.063  0.00362 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.2621 on 8 degrees of freedom
## Multiple R-squared:  0.9906, Adjusted R-squared:  0.9813
## F-statistic: 105.9 on 8 and 8 DF,  p-value: 2.622e-07
```

**Observation:** By observing the summary output of the multiple linear regression model, we inspect the `Estimate` and `Pr(>|t|)` (p-value) columns for each predictor. Iron (`Fe`) demonstrates the highest coefficient estimate magnitude among the metallic oxides for this specific subset of vehicle window glass. Based on this observation of the effect size (estimate), `Fe` is selected as the primary predictor for the Refractive Index in the subsequent steps.

## (b)

Run a simple linear regression of RI on the best predictor chosen in (a).

```
model_b = lm(RI ~ Fe, data = df_veh)

summary(model_b)

##
## Call:
## lm(formula = RI ~ Fe, data = df_veh)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2324 -1.0693 -0.2715  0.2907  3.7707
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.5007     0.4861  -1.030   0.3193
## Fe            8.1362     4.0780   1.995   0.0645 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.759 on 15 degrees of freedom
## Multiple R-squared:  0.2097, Adjusted R-squared:  0.157
## F-statistic: 3.981 on 1 and 15 DF,  p-value: 0.06452
```

### (c)

Can you further improve the regression of the refractive index of "Vehicle Window glass" on the predictor chosen by you in part (a)?

```
model_c = lm(RI ~ Fe + I(Fe^2), data = df_veh)
summary(model_c)

##
## Call:
## lm(formula = RI ~ Fe + I(Fe^2), data = df_veh)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6215 -1.1715 -0.1345  0.5985  3.5485
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.2785     0.4712  -0.591    0.564
## Fe          -12.1810    12.0408  -1.012    0.329
## I(Fe^2)      65.9600    37.0798   1.779    0.097 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.645 on 14 degrees of freedom
## Multiple R-squared:  0.3554, Adjusted R-squared:  0.2633
## F-statistic:  3.86 on 2 and 14 DF,  p-value: 0.04623
```

**Comparison**

```
stargazer(model_b, model_c, type = "text", title = "Comparison of Linear and
Non-Linear Models", column.labels = c("Linear Model", "Non-Linear Model"),
model.names = FALSE)

##
## Comparison of Linear and Non-Linear Models
## ===============================================================
##                          Dependent variable:
##              -------------------------------------------
##                                    RI
##                    Linear Model      Non-Linear Model
##                        (1)                 (2)
## ---------------------------------------------------------------
## Fe                    8.136*             -12.181
##                      (4.078)            (12.041)
##
## I(Fe2)                                   65.960*
##                                         (37.080)
##
## Constant              -0.501             -0.278
##                      (0.486)            (0.471)
##
## ---------------------------------------------------------------
## Observations           17                  17
## R2                   0.210               0.355
## Adjusted R2          0.157               0.263
## Residual Std. Error  1.759 (df = 15)     1.645 (df = 14)
## F Statistic        3.981* (df = 1; 15) 3.860** (df = 2; 14)
## ===============================================================
## Note:                      *p<0.1; **p<0.05; ***p<0.01
```

**Observation:** * By including the quadratic term `I(Fe^2)` in the regression model, we can capture potential non-linear relationships between `Fe` and `RI`.

- As the quadratic term is significant and the adjusted R-squared increases compared to the linear model, it suggests that the non-linear regression provides a better explanation of the relationship between `Fe` and `RI`.