Analysis

Carson Slater

2024-01-11

# Carson Slater, FIS Training Data

## ── Attaching packages ────────────────────────────────────── tidymodels 1.1.1 ──

## ✔ broom 1.0.5 ✔ recipes 1.0.8  
## ✔ dials 1.2.0 ✔ rsample 1.2.0  
## ✔ dplyr 1.1.4 ✔ tibble 3.2.1  
## ✔ ggplot2 3.4.4 ✔ tidyr 1.3.0  
## ✔ infer 1.0.5 ✔ tune 1.1.2  
## ✔ modeldata 1.2.0 ✔ workflows 1.1.3  
## ✔ parsnip 1.1.1 ✔ workflowsets 1.0.1  
## ✔ purrr 1.0.2 ✔ yardstick 1.2.0

## ── Conflicts ───────────────────────────────────────── tidymodels\_conflicts() ──  
## ✖ purrr::discard() masks scales::discard()  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ recipes::step() masks stats::step()  
## • Learn how to get started at https://www.tidymodels.org/start/

##   
## Please cite as:

## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

## Registered S3 method overwritten by 'printr':  
## method from   
## knit\_print.data.frame rmarkdown

# EDA

### Data Prep

### Plotting

# Modeling

## parsnip model object  
##   
##   
## Call:  
## stats::lm(formula = scaled\_psd ~ input\_psd + distance + distance\_sq,   
## data = data)  
##   
## Coefficients:  
## (Intercept) input\_psd distance distance\_sq   
## -1.012e+02 9.909e-01 -3.084e-02 1.103e-05

## parsnip model object  
##   
##   
## Call:  
## stats::lm(formula = scaled\_psd ~ input\_psd + distance + distance\_sq +   
## distance\_cb, data = data)  
##   
## Coefficients:  
## (Intercept) input\_psd distance distance\_sq distance\_cb   
## -9.705e+01 9.989e-01 -5.109e-02 4.037e-05 -1.224e-08

## parsnip model object  
##   
##   
## Call:  
## stats::lm(formula = scaled\_psd ~ input\_psd + input\_psd\_sq + distance +   
## distance\_sq, data = data)  
##   
## Coefficients:  
## (Intercept) input\_psd input\_psd\_sq distance distance\_sq   
## -1.028e+02 9.339e-01 -5.029e-04 -3.083e-02 1.102e-05

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | -101.1999105 | 0.0236274 | -4283.1519 | 0 |
| input\_psd | 0.9909187 | 0.0003419 | 2898.6110 | 0 |
| distance | -0.0308387 | 0.0000338 | -911.4964 | 0 |
| distance\_sq | 0.0000110 | 0.0000000 | 526.5981 | 0 |

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | -97.0488439 | 0.0158408 | -6126.5026 | 0 |
| input\_psd | 0.9989311 | 0.0001820 | 5489.6137 | 0 |
| distance | -0.0510868 | 0.0000507 | -1008.1293 | 0 |
| distance\_sq | 0.0000404 | 0.0000001 | 580.1965 | 0 |
| distance\_cb | 0.0000000 | 0.0000000 | -427.1462 | 0 |

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | -102.7879161 | 0.1456728 | -705.60804 | 0 |
| input\_psd | 0.9339271 | 0.0051702 | 180.63633 | 0 |
| input\_psd\_sq | -0.0005029 | 0.0000455 | -11.04721 | 0 |
| distance | -0.0308253 | 0.0000338 | -911.28976 | 0 |
| distance\_sq | 0.0000110 | 0.0000000 | 526.44031 | 0 |

### Model Metrics

| .metric | .estimator | .estimate |
| --- | --- | --- |
| mape | standard | 0.2802016 |

| .metric | .estimator | .estimate |
| --- | --- | --- |
| mape | standard | 0.1424803 |

| .metric | .estimator | .estimate |
| --- | --- | --- |
| mape | standard | 0.2796333 |

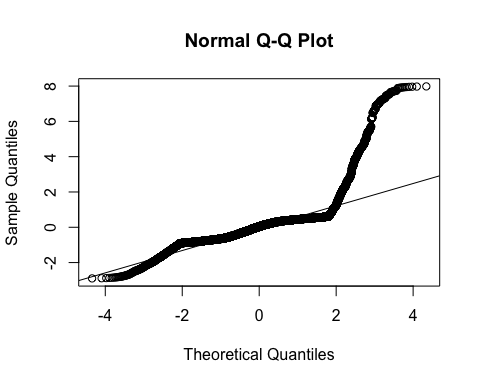
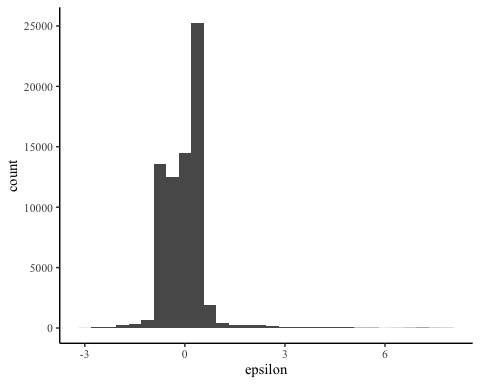
| .metric | .estimator | .estimate |
| --- | --- | --- |
| rmse | standard | 0.7233607 |

| .metric | .estimator | .estimate |
| --- | --- | --- |
| rsq | standard | 0.9937777 |

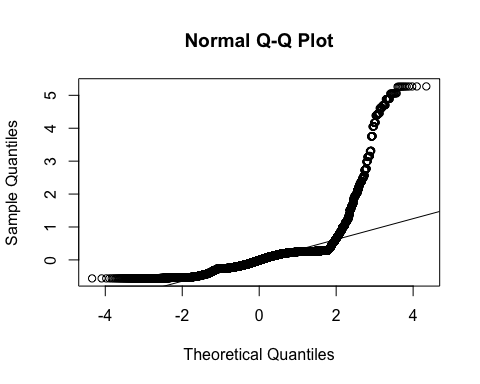
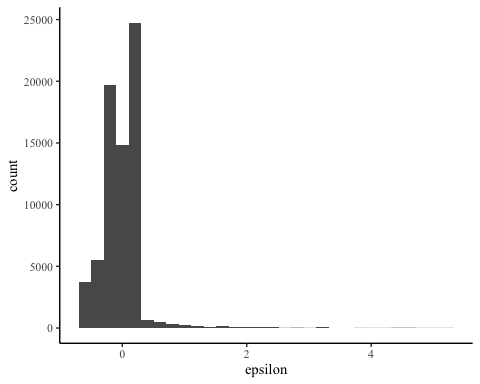
| .metric | .estimator | .estimate |
| --- | --- | --- |
| rsq | standard | 0.9982558 |

| .metric | .estimator | .estimate |
| --- | --- | --- |
| rmse | standard | 0.3829814 |

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



# Model Plotting

# Tables

aggregate\_data2 <- aggregate\_data |> mutate(  
 `Scaled PSD` = scaled\_psd,  
 `Input PSD` = input\_psd,  
 `Distance` = distance,  
 `Distance Sq.` = distance\_sq,  
 `Distance Cb.` = distance\_cb  
)  
  
model1 <- lm(`Scaled PSD` ~ `Input PSD` + `Distance` + distance\_sq,  
 data = aggregate\_data2)  
(model2 <- lm(`Scaled PSD` ~ `Input PSD` + `Distance` + `Distance Sq.` + `Distance Cb.`,  
 data = aggregate\_data2))

##   
## Call:  
## lm(formula = `Scaled PSD` ~ `Input PSD` + Distance + `Distance Sq.` +   
## `Distance Cb.`, data = aggregate\_data2)  
##   
## Coefficients:  
## (Intercept) `Input PSD` Distance `Distance Sq.` `Distance Cb.`   
## -9.705e+01 9.989e-01 -5.109e-02 4.037e-05 -1.224e-08

tidymodel1 <- model1 |> tidy()  
tidymodel2 <- model2 |> tidy()  
  
meticsmodel1 <- glance(model1)  
meticsmodel2 <- glance(model2)  
  
tidymodel2 |> kable()

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | -97.0488439 | 0.0158408 | -6126.5026 | 0 |
| Input PSD | 0.9989311 | 0.0001820 | 5489.6137 | 0 |
| Distance | -0.0510868 | 0.0000507 | -1008.1293 | 0 |
| Distance Sq. | 0.0000404 | 0.0000001 | 580.1965 | 0 |
| Distance Cb. | 0.0000000 | 0.0000000 | -427.1462 | 0 |

meticsmodel2 |> kable()

| r.squared | adj.r.squared | sigma | statistic | p.value | df | logLik | AIC | BIC | deviance | df.residual | nobs |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0.9982558 | 0.9982557 | 0.3829949 | 10168126 | 0 | 4 | -32633.19 | 65278.38 | 65333.41 | 10424.17 | 71065 | 71070 |

# stargazer(model1, model2,  
# title = "Regression table with stargazer",  
# label="tab2",  
# table.placement = "H",  
# column.labels = c("M1", "M2"),   
# model.numbers = FALSE,  
# header=FALSE)

# Bench Marking