EFFECTS OF SUPERPLASTICIZER ON CONCRETE COMPRESSIVE STRENGTHS AT 28 DAYS

Meng Ea
Email: s3716021@student.rmit.edu.au

Contents

1. Introduction

Concrete is one of the most commonly used construction materials in building structures in civil engineering. For many years, there have been many research projects conducted. These aims to investigate how different combinations of concrete mixture components and additions of new ones such as minerals and superplasticizers can improve concrete's properties.

The goal of this project is to develop a model that answers the research question: "What is the relationship between the amount of superplasticizer in the concrete mix to the concrete compressive strengths at 28 days while adjusting for potential confounders?"

2. Method

2.1. Data Description

Data used to answer the research question was collected from several experimental data of concrete mix samples (Yeh, 1998), where there is a total of 8 quantitative independent variables: cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, fine aggregate and age. The dependent variable is the concrete compressive strength, which is also quantitative. There are a total of 1030 rows of observations. The followings are ranges of values for each variable:

- Cement: 281.168 kg/m^3 to 540 kg/m^3
- Blast Furnace Slag: 0.000 kg/m^3 to 359.400 kg/m^3
- Fly Ash: 0.000 kg/m^3 to 200.100 kg/m^3
- Water: 121.800 kg/m^3 to 247.000 kg/m^3
- Superplasticizer: 0.000 kg/m^3 to 32.2000 kg/m^3
- Coarse Aggregates: 801.000 kg/m^3 to 1145.000 kg/m^3
- Fine Aggregates: 594.000 kg/m^3 to 992.600 kg/m^3
- Age: 1 day to 365 days
- Compressive Strengths: 2.330 MPa to 82.600 MPa

To answer the research question where only compressive strength at 28 days is concerned, the data was subsetted to include those observations at 28 days only. The new total of observations is, therefore, 425 samples of concrete mix.

2.2. Analytic Approach

The primary outcome was compressive strength, measured in MPa. The primary explanatory variable was superplasticizer content in kilograms per cubic metre of concrete mix (kg/m^3).

First, the full model, including superplasticizer and potential confounders as well as effect modifications, was specified based on general knowledge and exploratory data analysis. A confounder is defined as meeting the following criteria:

- A variable that is associated with the primary explanatory variable, superplasticizer.
- A variable that is a predictor of the outcome variable, compressive strength.
- However, one that is not on the causal pathway between explanatory and outcome variable.

All potential confounders were evaluated as follows:

- Cement and water are associated with superplasticizer because in practice, to determine the amount of superplasticizer, a marsh cone test is conducted on cement paste that is composed of cement and water to obtain desirable concrete workability. Further, fly ash and blast furnace slag are common mineral additions to replace some part of cement content in certain concrete mix. Therefore, they are also associated with superplasticizer.
- Cement, water, fly ash and blast furnace slag has long been proven through both empirical research and studies of involved mechanisms that they have effects on concrete compressive strength. Consequently, they are also predictors of the concrete compressive strength.
- However, water is on the causal pathway. This is because superplasticizer acts as water reducer to the concrete mix and this in turns cause compressive strength to increase.
- Coarse aggregates and fine aggregates have been shown to have negligible effects on compressive strengths when the cement and other mineral additions are taken into account. This does not meet the criteria of being predictors of compressive strength and thus are not confounders.
- As a result, only cement, fly ash and blast furnace slag are potential confounders.

Potential effect modifications were observed from exploratory data analysis of the dataset, where compressive strength was plotted against superplasticizer content at different levels of cement, fly ash and blast furnace slag as illustrated in Figure 1, 2 and 3. From these figures, it is seen that the relationship between compressive strength and superplasticizer could depend on different levels of cement, fly ash and blast furnace slag. In other words, it is expected that the compressive strength increases or decreases with superplasticizer content, but the rate of increase or decrease should be different for different levels of cement, fly ash and blast furnace slag.

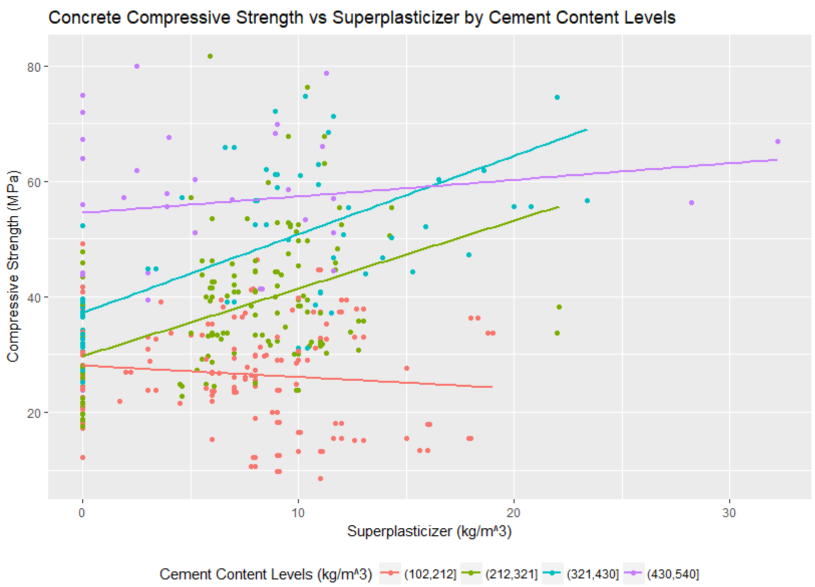

Figure 1 Concrete Compressive Strength vs Superplasticizer by Cement Content Levels

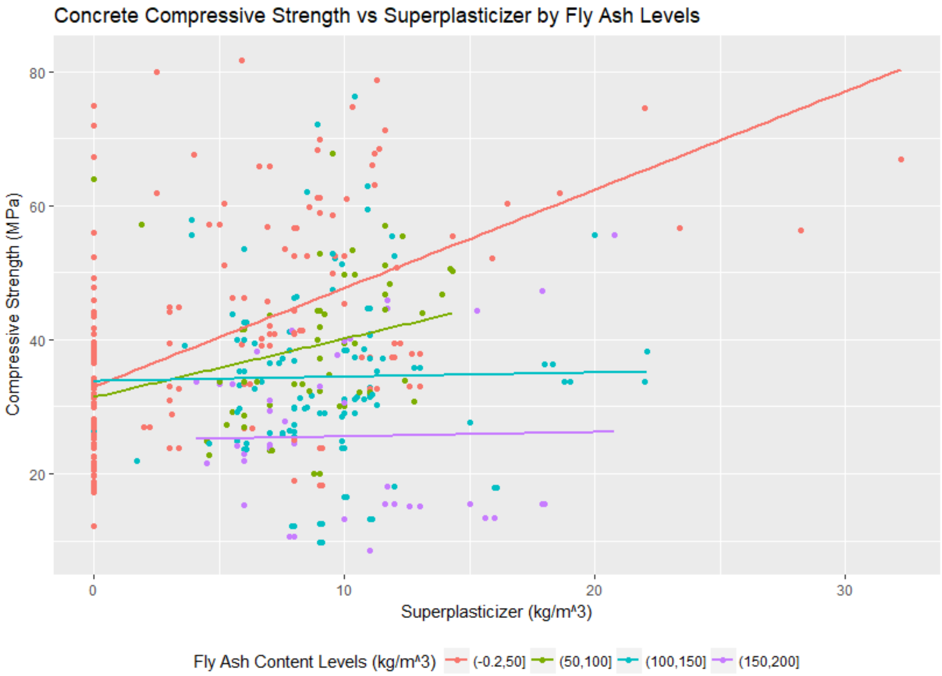

Figure 2 Concrete Compressive Strength vs Superplasticizer by Fly Ash Content Levels

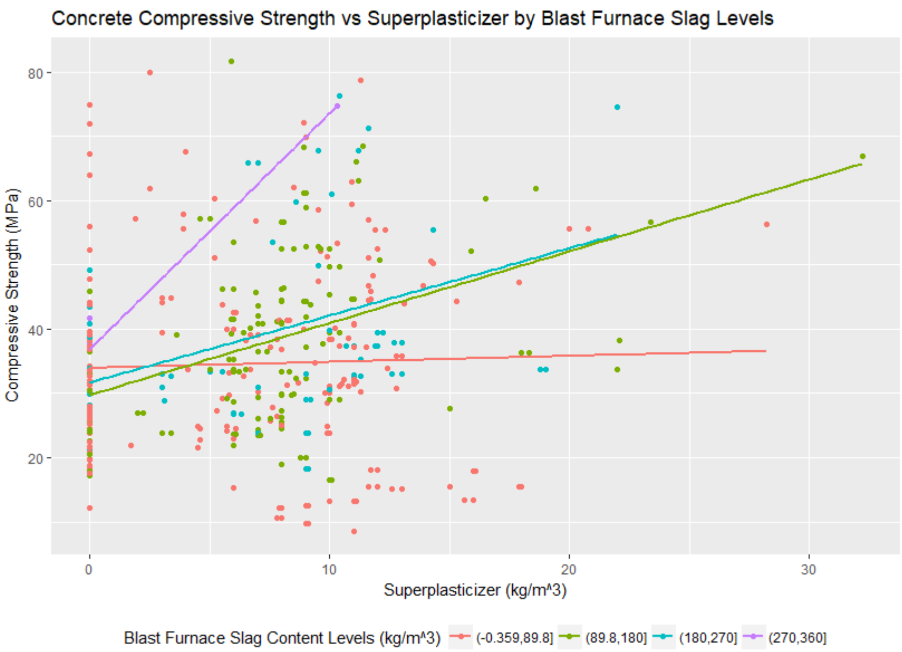

Figure 3 Concrete Compressive Strength vs Superplasticizer by Blast Furnace Slag Content Levels

The full model specified below was used to fit the data using multiple linear regression method:

Compressive Strength = Superplasticizer + Cement + Fly Ash + Blast Furnace Slag + Super:Cement + Super:Fly Ash + Super:Blast Furnace Slag.

Here, the full model was checked for model adequacy against multiple linear regression assumptions. Those assumptions are for the error term:

- The errors have constant variance.

- The errors should be independent of each other.
- The errors are normally distributed.

The tests used to check these are:

- Non-constant Variance Test
  - H_o: the errors have constant variance.
  - H_a: the error variance changes with the level of the fitted values.
- Durbin-Watson Test for Autocorrelated Errors
  - H_o: the errors are not autocorrelated.
  - $H_{a:}$ the errors are autocorrelated.
- Shapiro-Wilk Test
  - H_o: errors are normally distributed.
  - H_a: errors are not normally distributed.

All analyses were conducted using Rstudio for Windows. All tests were two-tailed and the significance level was set at alpha = 0.05.

The model was also checked for multicollinearity by inspecting the variance inflation factor. Variance inflation factor of less than 5 is satisfactory, or in other words, there is no significant multicollinearity.

All possible regressions were then performed, and the final three models were chosen based on the minimum Mallow's C_p Statistic. This metric was chosen because this is not a predictive model but rather a comparative hypothesis studying the effect of a primary explanatory variable on primary response outcome where variance in the fitted values of response outcome is most desirable. The models must also contain the superplasticizer term as it is a primary explanatory variable being investigated.

After that, each of the final three models were then checked for model adequacy and multicollinearity similarly to the full model. Finally, they will be discussed and evaluated to choose the best model possible.

3. Results

Summary statistics of all variables is shown in the following page:

Table 1 Summary Statistics of All Variables

Variables	Minimum	First Quartile	Median	Mean	Third Quartile	Maximum
Compressive Strength (Mpa)	8.54	26.23	33.76	36.75	44.39	81.75
Superplasticizer Content (kg/m^3)	0	0	7.8	6.996	10.3	32.2
cement content (kg/m^3)	102	160.2	261	265.4	323.7	540
Blast Furnace Slag Content (kg/m^3)	0	0	94.7	86.29	160.5	359.4
Fly Ash Content (kg/m^3)	0	0	60	62.8	120	200
Water Content (kg/m^3)	121.8	171	185	183.1	193.3	247
Coarse Aggregates Content (kg/m^3)	801	882.6	953.2	956.1	1013.2	1145
Fine Aggregates Content (kg/m^3)	594	712	769.3	764.4	811.5	992.6

The full model initially failed all tests for assumptions. This was rectified by performing the followings:

- As interaction terms were added, all independent variables were centred by subtracting each value by their corresponding means to avoid multicollinearity.
- Further, compressive strength was transformed into the square root of compressive strength to eliminate non-constant variance among residuals.
- In addition to this, the Cochrane-Orcutt method was used to eliminate the autocorrelations of residuals.

When the hypothesis tests for assumptions were conducted again, the model was able to meet the assumptions for non-constant error variance and uncorrelated errors as well as having no significant multicollinearity. Other attempts at transformations and standardising or scaling, however, were unsuccessful in enabling the model to meet normally-distributed errors assumption.

The all possible regression was performed on the full model regardless of unable to meet the criteria to generate the top three models based on C_p mallow statistics. The three models outlined below are in order of having the smallest C_p to bigger C_p:

- Model 1: *Compressive Strength ~ Superplasticizer + Cement + Superplasticizer:Fly Ash + Superplasticizer:Blast Furnace Slag.*

- Model 2: *Compressive Strength ~ Superplasticizer + Superplastsicizer:Fly Ash + Superplasticizer:Blast Furnace Slag.*
- Model 3: *Compressive Strength ~ Superplasticizer + Blast Furnace Slag + Superplasticizer:Fly Ash + Superplasticizer:Blasts Furnace Slag.*

When each model was checked for model adequacy, similarly to the full model, they only met the assumptions of non-constant variance, uncorrelated errors and no significant multicollinearity, but not for normally-distributed errors. The terms that are significant in all models are superplasticizer and intercepts. Interaction terms are statistically significant, but coefficients are small and uninterpretable in the context of measurement units.

Table 2 illustrates a summary of the relevant results of the three models.

Table 2 Summary of Model 1, 2 and 3 Intercept and Superplasticizer Coefficient

Final Models	Intercept	Coefficients for Superplasticizer	The standard error of Coefficient for Superplasticizer
Model 1	3.993	0.023	0.010
Model 2	3.998	0.022	0.010
Model 3	4.009	0.021	0.010

```
Call:
lm(formula = compressive ~ super + cement + super:flyash + super:slag,
    data = concrete)

Residuals:
     Min       1Q   Median       3Q      Max
-3.08530 -0.60163  0.08743  0.73404  2.86992

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.9931142  0.0586703  68.060   <2e-16 ***
super         0.0229311  0.0100540   2.281   0.0231 *
cement        0.0009325  0.0005087   1.833   0.0675 .
super:flyash -0.0003397  0.0001682  -2.019   0.0441 *
super:slag   -0.0001974  0.0001105  -1.786   0.0748 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.06 on 419 degrees of freedom
Multiple R-squared:  0.04189,  Adjusted R-squared:  0.03275
F-statistic:  4.58 on 4 and 419 DF,  p-value: 0.001248
```

Figure 4 Model 1 Summary

4. Discussion

Because the purpose of this research is to investigate the effect of superplasticizer content on concrete compressive strength, smaller standard error of coefficient for superplasticizer term is more desirable as the least biased estimate is sought. According to Table 2, these are the same for all models. Further, all models pass the same criterion for model adequacy check and have insignificant multicollinearity, except for non-normality of errors.

As a result, C_p statistic was chosen as an evaluation metric for choosing the best model. In this case, it was Model 1. Another reason is that this model has adjusted for cement content, which must present in all concrete mix. This is seen as more realistic and meaning in the intended operating environment.

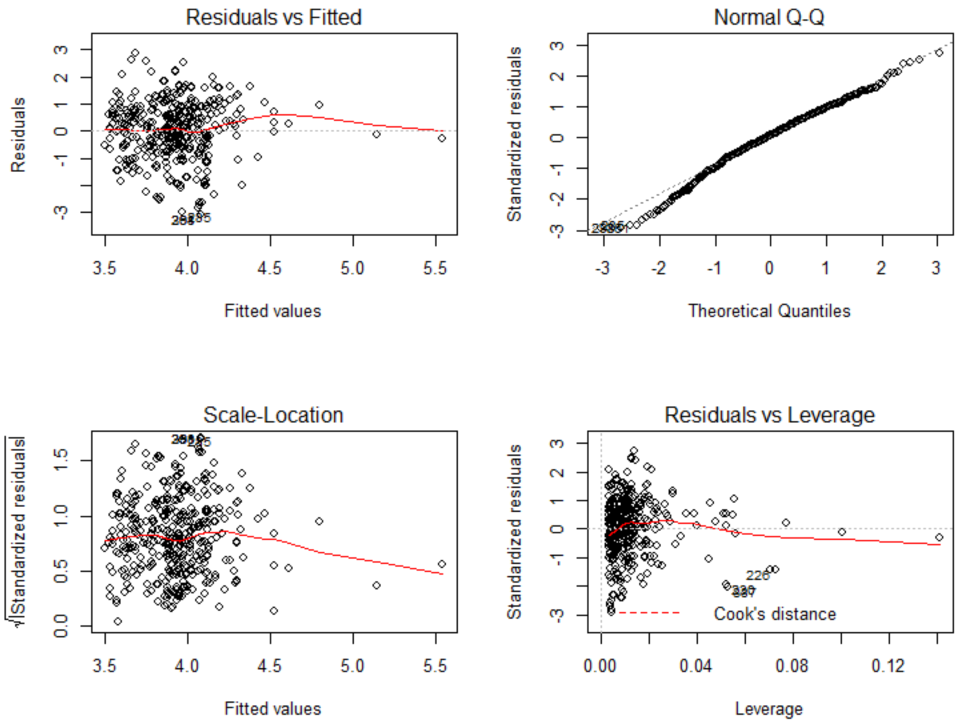

Figure 5 Residual Plots of Model 1

According to normal Q-Q plot in figure 4, although the model failed the test, there is no gross non-normality shown. Non-normality is more serious when the t or F statistics and confidence and prediction intervals needed. At this stage, only the point estimate can be confidently interpreted as the estimate is still unbiased according to ordinary least square properties of multiple linear regression.

According to Table 2, the intercept and superplasticizer coefficient can be interpreted as followings:

- $(3.993)^2$ MPa $\approx$ 16MPa reflects the average concrete compressive strength at 28 days for the mix with an average superplasticizer content of 6.996 kg/m^3 and average cement content of 265.4 kg/m^3.
- For every increase of 1 kg/m^3 in superplasticizer content in the concrete mix, there is an increase of $(0.023)^2 = 0.000524$ MPa in the mean of concrete compressive strength adjusting for cement content.

5. Conclusion

In conclusion, the final model answered the research question that there is a positive relationship between the amount of superplasticizer and concrete compressive strength at 28 days when adjusting for cement content. The model was useful for the point estimate of this relationship only. However, since the non-normality assumption could not be met at this stage, confidence intervals and p-value indicating statistical significant in hypothesis testing cannot be relied upon.

It is recommended to use maximum likelihood to estimate parameters for distribution of errors different from normal distribution or use robust statistical inference methods.

6. Reference

Yeh, I.-C 1998, ‘Modelling of strength of high-performance concrete using artificial neural networks’, *Cement and Concrete Research*, Vol. 28, No. 12, pp. 1797-1808.

Appendix

```
### import packages
library(dplyr)
library(tidyr)
library(ggplot2)
library(QuantPsyc)
library(car)
library(TSA)
library(olsrr)
library(dplyr)
library(stringr)
library(GGally)

#import data
concrete <- read.table("Concrete_Data.csv", sep = ",", col.names =
c("cement", "slag", "flyash", "water", "super", "coarseagg", "fineagg",
"age", "compressive"),header = FALSE, skip = 1)
concrete <- concrete %>% filter(age == 28)
head(concrete)
concrete <- concrete[,-8]
summary(concrete)
head(concrete)
dim(concrete)

#plot histogram of compressive strength
hist(concrete$compressive)

#plot response vs super
ggplot(concrete, aes(x = super, y = compressive)) + geom_point()

# explore interaction between super and cement, flyash, slag on compressive
strength
concrete$cement_binned <- cut(concrete$cement, 4)

ggplot(concrete, aes(x = super, y = compressive, color = cement_binned,
group = cement_binned)) + geom_point() + geom_smooth(method = "lm", se =
FALSE) + labs(colour = "Cement Content Levels (kg/m^3)", x =
"Superplasticizer (kg/m^3)", y = "Compressive Strength (MPa)", title =
```

```
"Concrete Compressive Strength vs Superplasticizer by Cement Content
Levels") + theme(legend.position = "bottom")

concrete$flyash_binned <-cut(concrete$flyash, 4)

ggplot(concrete, aes(x = super, y = compressive, color = flyash_binned,
group = flyash_binned)) + geom_point() + geom_smooth(method = "lm", se =
FALSE) + labs(colour = "Fly Ash Content Levels (kg/m^3)", x =
"Superplasticizer (kg/m^3)", y = "Compressive Strength (MPa)", title =
"Concrete Compressive Strength vs Superplasticizer by Fly Ash Levels") +
theme(legend.position = "bottom")

concrete$slag_binned <- cut(concrete$slag, 4)

ggplot(concrete, aes(x = super, y = compressive, color = slag_binned, group
= slag_binned)) + geom_point() + geom_smooth(method = "lm", se = FALSE) +
labs(colour = "Blast Furnace Slag Content Levels (kg/m^3)", x =
"Superplasticizer (kg/m^3)", y = "Compressive Strength (MPa)", title =
"Concrete Compressive Strength vs Superplasticizer by Blast Furnace Slag
Levels") + theme(legend.position = "bottom")

# try z-normalisation
concrete[, c(1,2,3,4,5,6,7)] <- scale(concrete[, c(1,2,3,4,5,6,7)], center
= TRUE, scale = FALSE)
head(concrete)
class(concrete)
concrete <- as.data.frame(concrete)

# try transforming response variable
concrete$compressive = concrete$compressive^(1/2)
hist(concrete$compressive)
head(concrete)

# specifying model_1
model_1 <- lm(compressive ~ super + cement + flyash + slag + super:cement +
super:flyash + super:slag, data = concrete)
summary(model_1)

#residual plots
plot(model_1)

#checking multicollinearity
vif(model_1)

# checking constant variance
ncvTest(model_1)

# checking autocorrelation
acf(model_1$residuals)
durbinWatsonTest(model_1)

#test for normally distributed errors
shapiro.test(model_1$residuals)
hist(concrete$compressive)

# try cochrane-orcutt method
ncvTest(model_1)
acf(model_1$residuals)
durbinWatsonTest(model_1)
p = durbinWatsonTest(model_1)$r
print(p)
```

```
head(concrete)
for (i in 2:length(concrete$compressive))
  concrete$compressive[i-1] = concrete$compressive[i]
p*concrete$compressive[i-1]
  concrete$cement[i-1] = concrete$cement[i] - p*concrete$cement[i-1]
  concrete$water[i-1] = concrete$water[i] - p*concrete$water[i-1]
  concrete$super[i-1] = concrete$super[i] - p*concrete$super[i-1]
  concrete$flyash[i-1] = concrete$flyash[i] - p*concrete$flyash[i-1]
  concrete$slag[i-1] = concrete$slag[i] - p*concrete$slag[i-1]

  concrete <- concrete[-425,]

# run the model again after cochrane-orcutt method
model_1 = lm(compressive ~ super + cement + flyash + slag + super:cement +
super:flyash + super:slag, data = concrete)
summary(model_1)

#checking multicollinearity again
vif(model_1)

# checking constant variance again
ncvTest(model_1)

# checking autocorrelation again
acf(model_1$residuals)
durbinWatsonTest(model_1)

#test for normally distributed errors again
shapiro.test(model_1$residuals)
hist(concrete$compressive)

#residual plots
model_1_aug <- augment(model_1)
head(model_1_aug)
ggplot(model_1_aug, aes(x = .resid^2, y = .fitted)) + geom_point()
concrete_diag <- cbind(concrete, model_1_aug)
qqPlot(model_1$residuals, dist = "norm")
qplot(concrete_diag$.resid, geom = "histogram")

# plotting distribution of variables
qplot(concrete$compressive, geom = "histogram")
qplot(concrete$cement, geom = "histogram")
qplot(concrete$super, geom = "histogram")
qplot(concrete$flyash, geom = "histogram")
qplot(concrete$slagh, geom = "histogram")

#all possible regressions
all_models<-ols_step_all_possible(model_1)
all_models
str(all_models)
all_models$predictors
plot(all_models)

# choosing models that contain super term
final_models <-  all_models[, c(3, 7, 8, 11)] %>%
filter(str_detect(predictors, "\\bsuper\\b(?!:)"))
str(final_models)
colnames(final_models)

# arrange the top 8 models based on cp statistics
```

```
final_8 <- head(final_models %>% arrange(cp), 8)
final_8$p <- sapply(strsplit(final_8$predictors, " "), length)
final_8

# final three models chosen are
final_model_1 <- lm(compressive ~ super + cement + super:flyash +
super:slag, data = concrete)
ols_regress(compressive ~ super + cement + super:flyash + super:slag, data
= concrete)

final_model_2 <- lm(compressive ~ super + super:flyash + super:slag, data =
concrete)
ols_regress(compressive ~ super + super:flyash + super:slag, data =
concrete)

final_model_3 <- lm(compressive ~ super + slag + super:flyash + super:slag,
data = concrete)
ols_regress(compressive ~ super + slag + super:flyash + super:slag, data =
concrete)

#residual analysis of final_model_1
par(mfrow = c(2,2))
plot(final_model_1)

  #checking multicollinearity
  vif(final_model_1)

  # checking constant variance
  ncvTest(final_model_1)

  # checking autocorrelation
  acf(final_model_1$residuals)
  durbinWatsonTest(final_model_1)

  #test for normally distributed errors
  shapiro.test(final_model_1$residuals)
  hist(concrete$compressive)

#residual analysis of final_model_2
plot(final_model_2)

  #checking multicollinearity
  vif(final_model_2)

  # checking constant variance
  ncvTest(final_model_2)

  # checking autocorrelation
  acf(final_model_2$residuals)
  durbinWatsonTest(final_model_2)

  #test for normally distributed errors
  shapiro.test(final_model_2$residuals)
  hist(concrete$compressive)

#residual analysis of final_model_3
plot(final_model_3)

  #checking multicollinearity
  vif(final_model_3)
```

```
  #checking constant variance
  ncvTest(final_model_3)

  #checking autocorrelation
  acf(final_model_3$residuals)
  durbinWatsonTest(final_model_3)

  #test for normally distributed errors
  shapiro.test(final_model_3$residuals)
  hist(concrete$compressive)

# summary of the best model
summary(final_model_1)
```