Predicting Concrete Strength

Javier Rondon, Daniele Grandi

4/11/2022

Contents

1 Introduction

Concrete is the second most-used substance in the world, after water[1]. In the construction industry, concrete is the most used building material, and is found in a wide range of structures, such as buildings, bridges, roads, and dams.

Concrete is a composite material comprised of aggregate solids glued together with a paste that hardens over time. Concrete has three essential ingredients: cement, water, and solid aggregates[2]. Cement and water form a pasted called the binder. Inert coarse materials such as coarse rocks or gravel form the filler. The binder paste glues the solids in the filler into a conglomerate.

When the cement is in contact with water, it undergoes a hydration process. In this reaction, calcium silicates are produced. Equation 1 shows this reaction for tricalcium silicate. This molecule determines the strength of the concrete at the early stages of curing[3].

$$2Ca_3SiO_5 \rightarrow 3CaO.2SiO_2.4H_2O + 3Ca(OH)_2 + 173.6kJ \tag{1}$$
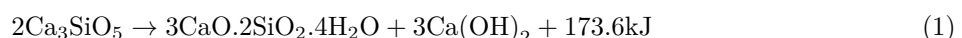
The properties of the cement paste determine the properties of the concrete. Over time, the concrete will continue to harden and gain strength for a long period of time.

The filler or aggregates are the solids glued together by the cement paste. They range in size and shape, from sand to coarse rocks. These aggregates can represent 70 to 80 percent of the volume of concrete.

In addition to these essential ingredients, modern concrete mixtures also incorporate additional ingredients, such as fly ash, blast furnace slag, and other chemical additives to reduce costs, accelerate the curing time or improve workability and increase strength[4].

The water/cement ratio (WC ratio) is considered the most important characteristic of concrete[5]. This property controls the strength, workability and durability of the product. There is a large body of research showing the importance of this property for the use of concrete[6]

The design of concrete mixtures depends on the desired characteristics of the concrete for a given application, site requirements and cost considerations[7]. Our team is interested in learning **how the water/cement ratio affects the strength of concrete**.

Table 1: A summary of the covariates used in the study, and their applied transformations.

Variable name	Common name	Unit of measure	Transformation
WC_ratio	Water to cement ratio	NA	Log
Age	Age	Days	Log
BFS	Blast furnace slag	NA	Categorical
Fly_Ash	Fly ash	NA	Categorical
SP	Superplasticizer	NA	Categorical

---

[1] Gagg, Colin R. "Cement and concrete as an engineering material: An historic appraisal and case study analysis." Engineering Failure Analysis 40 (2014): 114-140.

[2] https://concretesupplyco.com/concrete-basics/

[3] http://matse1.matse.illinois.edu/concrete/prin.html

[4] https://www.ijser.org/researchpaper/Use-of-Blast-Furnace-Slag-Aggregate-in-Concrete.pdf

[5] https://mastercivilengineer.com/significance-of-water-to-cement-ratio/

[6] Aitcin,P.-C. "The importance of the water-cement and water-binder ratios." Science of Technology of Concrete Admixtures (2016): 3-23 https://www.sciencedirect.com/science/article/pii/B9780081006931000011

[7] https://concretesupplyco.com/concrete-mix-design/

2   Data and Methodology

2.1   Data Source

Our team sourced a multivariate dataset from the UC Irvine Machine Learning Repository[8] describing the ratio of ingredients and age of 1030 different concrete mixtures, and their measured compressive strength. The dataset is built from 17 different sources, which brings some limitations around specific control of the sub-type of ingredients across samples (i.e.: the type of ash was not controlled, just the amount). Nonetheless, all ingredient ratios were controlled within reasonable limits.

In our research design, we randomly sampled 30% of the observations to keep aside for exploration, while the remaining 70% of the observations were used to fitting the models.
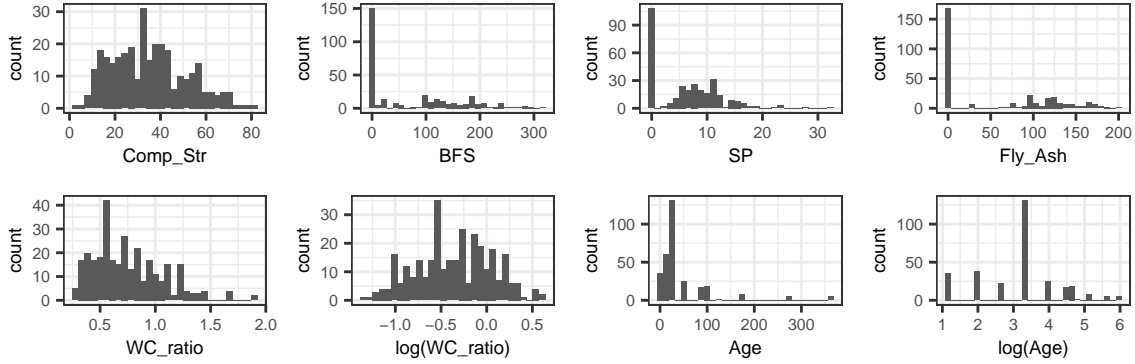

Figure 1: Distribution of covariates and selected transformations.

2.2   Characteristics and transformations

The `compressive strength` measurement is our outcome variable. `Compressive strength` is measured in megapascals (MPa), and is observed in the data from 3.32 MPa to 81.75 MPa, with a mean of 35.07 MPa. Figure 1 shows the distribution of `compressive strength`, which is observed to be relatively normal.

The `WC ratio` was calculated from the `water` and `cement` variables in the dataset. The `WC ratio` is commonly expressed as a mass ratio since the components of concrete are weighed before preparation.

The remaining variables in the dataset correspond to the age of the concrete, the aggregates in the filler and additives. These variables have an effect on the `compressive strength` of the final mixture, as explained in the introduction, and will serve as covariates. Table 1 provides a summary of the variable names in the dataset, their common name, their unit of measure, and whether the variable will be considered metric or categorical. Figure 1 shows the distribution of the covariates, as well as relevant transformations.

Figure 2 shows the correlations between our variables of interest. We observe a lack of perfect correlation between measured variables.

In Figure 3, the left panel shows `compressive strength` as a function of `w/c ratio`. We notice a downward trend of the sample's `compressive strength` as the `WC ratio` increases. This indicates that the reducing the amount of water in the mixture increases the concrete strength.

In Figure 3, in the right panel, we see a positive trend between `age` and `compressive strength`. This indicates that as the concrete ages, its strength increases. first rapidly and then slowly. The above trends warrant further investigation, and particular care will be taken to observe these trends in our models, given their relationship to key performance metrics for our business.
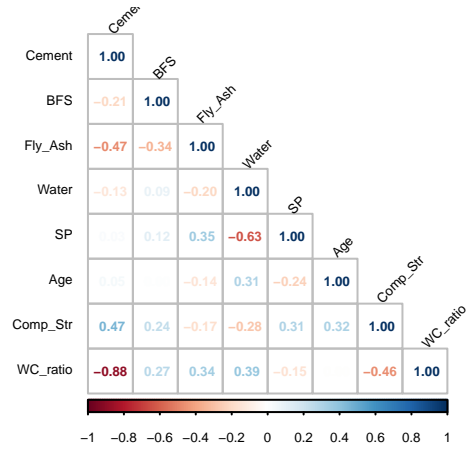
---

[8]https://archive-beta.ics.uci.edu/ml/datasets/concrete+compressive+strength

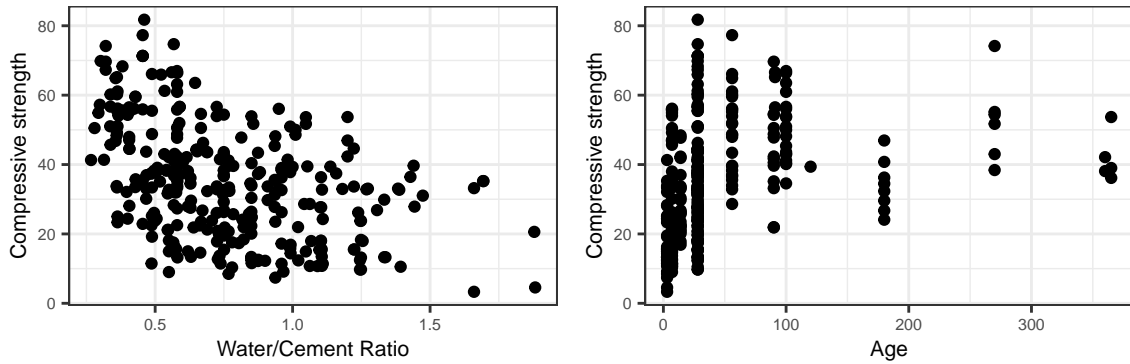Figure 2: Correlation table between variables in the dataset

Figure 3: Scatterplot showing a negative relationship between `w/c ratio` and `compressive strength` (left), and scatterplot showing a positive relationship between `age` and `compressive strength` (right)

2.3 Transformations

While there are no NA values in our dataset, we note the presence of several 0 values, which we will leverage to transform those variables to categorical.

From Figure 1, we observe that the `WC ratio` has a skewed distribution towards the lower range. We will apply a logarithmic transformation to this variable to alleviate this issue. Similarly, in the same figure, we can see from the distribution of `age` that most concrete samples were tested for less than 100 days, and only a few samples were tested for up to a year. We will apply a logarithmic transformation to this variable as well.

While all measured variables are metric, we will consider the presence of additives in the model as categorical variables. `Ash`, `BFS`, and `SP` were introduced as considered categorical (either included or excluded in the mixture), given that the dataset contains enough observations of both cases.

3 Proposed Models

The primary outcome that we are interested in is the `compressive strength` of a concrete sample as measured in megapascal (MPa). From the literature, we know that the strength of the final concrete block is primarily dependent on the `WC ratio` in the mixture, which we can model with Equation 2. Other important factors in concrete strength are `age` (Equation 3), as well as the use of other ingredients in the mixture: `slag`, `ash`, and `superplasticizer` (Equation 4). We are interested in assessing the importance of the `WC ratio` in the strength of the concrete as this is an easy-to-measure ratio during the production process that has a direct effect on the concrete strength, allowing us to exceed customer expectations.

$$Strength = \tilde{\beta}_0 + \tilde{\beta}_1 log(\frac{Water}{Cement}Ratio) + \epsilon \tag{2}$$

$$Strength = \tilde{\beta}_0 + \tilde{\beta}_1 log(\frac{Water}{Cement}Ratio) + \tilde{\beta}_2 log(Age) + \epsilon \tag{3}$$

$$Strength = \tilde{\beta}_0 + \tilde{\beta}_1 log(\frac{Water}{Cement}Ratio) + \tilde{\beta}_2 log(Age) + \tilde{\beta}_3 Slag + \tilde{\beta}_4 Ash +$$
$$\tilde{\beta}_5 Superplasticizer + \epsilon \tag{4}$$

4 Results

We fit our three models to the test data, and report our results in Table 2, which includes the coefficients, robust standard error, and indication of statistical significance of the covariates, and adjusted R^2 measures of each model.

4.1 Statistical Significance

From Table 2, we see that all coefficients are significant based on the standard t-test, except for the coefficient of the `constant` in Model 2.

The adjusted R^2 value provides an indication that adding additional covariates significantly increases the model fit. Nonetheless, to assess which model best fits our data, we can also look at the mean squared residual (MSR) values of each model. We find the MSR to be 194.0, 98.1, and 57.2 for models 1, 2, and 3, respectively. Moreover, to provide evidence of the additional parameters that we have estimated in models 2 and 3, we can conduct two F-test to evaluate which model is more appropriate. Both tests result in p-values of 2.2e-16, indicating that there is indeed strong statistical significance, and out of the three models, model 3 does indeed fit our data best.

Table 2: Model Estimates

	Dependent variable:		
	Compressive Strength		
	(1)	(2)	(3)
log(WC ratio)	−23.120*** (1.227)	−23.501*** (0.843)	−23.948*** (0.922)
log(age)		8.140*** (0.325)	8.233*** (0.240)
Slag			8.995*** (0.652)
Superplasticizer			9.267*** (1.135)
Ash			−2.354** (1.130)
Constant	27.385*** (0.621)	1.549 (1.021)	−8.633*** (0.849)
Observations	721	721	721
R^2	0.314	0.653	0.798
Adjusted R^2	0.313	0.652	0.796
Residual Std. Error	13.948 (df = 719)	9.925 (df = 718)	7.596 (df = 715)
F Statistic	329.227*** (df = 1; 719)	676.139*** (df = 2; 718)	563.924*** (df = 5; 715)

Note: *p<0.1; **p<0.05; ***p<0.01

4.2 Practical Significance

In order to evaluate the practical significance of our model, we can showcase how we might use the model in practice with two examples.

We are interested in answering the following question: "given a specific mixture, when can construction continue?" In Figure 4, we use our model to predict the compressive strength of concrete as a function of a range of WC ratios, from 0.3 to 0.8, which are typical practical ranges of concrete. If the mixture goes beyond these ranges, the resulting mixture will not harden into concrete; thus, we are particularly interested in how our model behaves in this range. From the plot, we can see that, as the days increase, the compressive strength of the concrete also increases steeply at first and then at a slower rate. This behavior matches the existing literature and our background knowledge. Moreover, the plot also shows that as the WC ratio increases, the concrete strength decreases, which is again as expected.
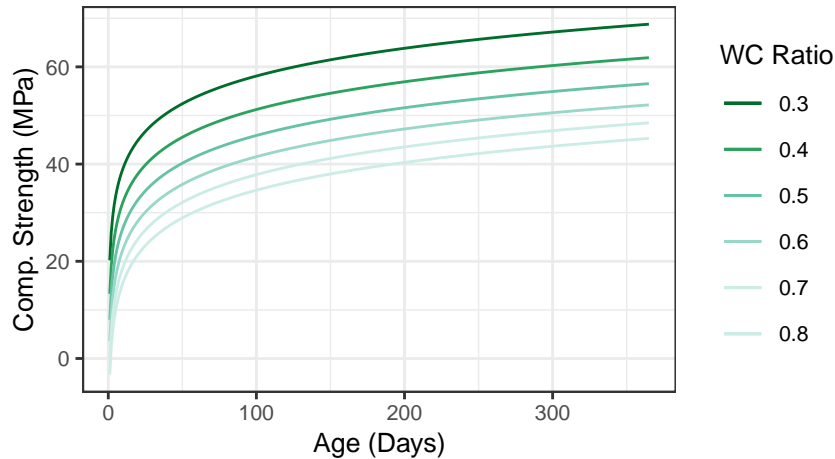

Figure 4: Relationship between age and strength as a function of WC ratio

Another question that we are interested in answering is: "how do additives affect compressive strength?" Figure 5 shows what happens when different aggregates are included in a base mixture. Our baseline is just composed of water and cement, and we can see from the plot that the addition of ash will reduce the

compressive strength for any given amount of drying time. This is expected, as ash is used to improve mixability, while the poor quality of the ash often results in weaker concrete. On the other hand, we can see from the plot that including slag and superplasticizer positively impacts the compressive strength. This again is expected and validated in recent literature.

This model will allow us to make a data-driven decision regarding the number of days before the concrete reaches a specific compressive strength, which is a valuable insight for our business, as construction delays are very costly.
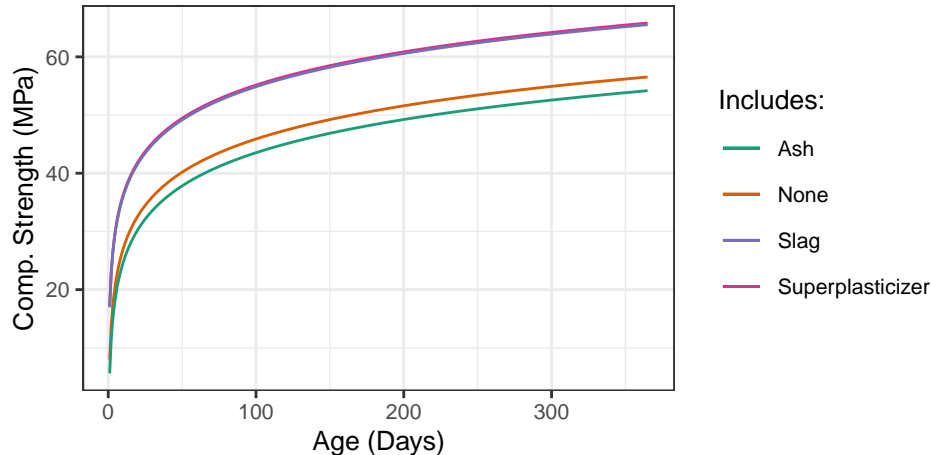

Figure 5: Relationship between age and strength when different aggregates are included in the mixture.

5 Model Limitations

5.1 Statistical Limitations

Next, we evaluate the three large sample model assumptions.

1. **IID Data:** To assess whether the data meets the IID assumption, we need to evaluate the sampling process. Unfortunately, the data was gathered from 17 different sources, and no specific documentation is provided regarding the sampling process. Nonetheless, the authors of the dataset state that "a determination was made to ensure that these mixtures were a fairly representative group governing all of the major parameters that influence the strength of high-performance concrete"[9]. Figure 1, showing the distribution of the variables, seems to indicate quasi-normal distributions that might have been created using a randomized approach in preparation for further experimentation. Nonetheless, this assumption remains the most concerning to our study. However, we believe that we have enough data from the 17 different sources to mitigate any issues that might arise from failure to meet the IID assumption on each source.

2. **No Perfect Colinearity:** We note from Figure 2 that no perfect colinearity exists between variables. Moreover, a BLP should exist as the distribution of `compressive strength` does not have heavy tails.

3. **Linear Conditional Expectation:** To assess whether there is a linear conditional expectation, we can observe the predicted values vs. the model's residuals, shown in Figure 6. From the plot, we can observe that the residuals of model 3 appear to be linear above 15 MPa of compressive strength. However, at lower values, the mean of the residuals starts to increase, indicating non-linear relationships.

---

[9]Yeh, I-C. "Modeling of the strength of high-performance concrete using artificial neural networks." Cement and Concrete Research 28.12 (1998): 1797-1808.

Nonetheless, given that our standard operating parameters are well below 15 MPa (typically above 40MPa), this behavior does not seem of concern.
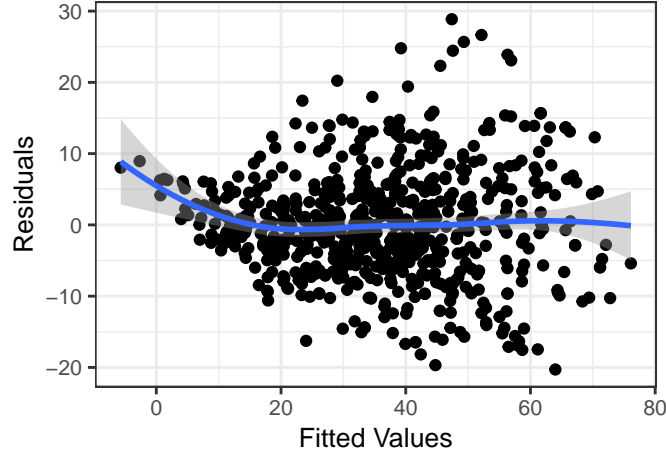

Figure 6: Residual of Model 3 plotted against fitted values

5.2 Structural limitations

5.2.1 Omitted variables

We have reviewed the literature and identified several omitted variables that have an effect on the compressive strength of concrete. Figure 7 shows these variables in a causal diagram.

We intentionally omitted the variables corresponding to the coarse and fine aggregates. We omitted using these variables because the aggregates' effect on the compressive strength is through the workability, which affects the amount of water used in the mixture.

The variables shown with the dotted line in Figure 7 were omitted from the model since we do not have data for them. We will discuss the impact of each omitted variable below.

Water quality: The presence of impurities in the water reduces the compressive strength of concrete. A reduction of the water's PH value will reduce the compressive strength of concrete.

Cement composition Cement composition significantly affects the hydration reaction and the properties of the binder. The direction of the effect on the compressive strength of concrete depends on the ingredients of the cement.

Environment temperature Temperature affects concrete strength. High temperatures may lead to cracks and reduce the workability of the concrete mixture.

Environment Humidity An increase in relative humidity will decrease the compressive strength of concrete.

Mixing technique The method selected to mix the concrete will impact the compressive strength. Therefore, the method must ensure that the concrete is homogeneous and consistent. The method must also consider the mixing time to ensure the ingredients blend correctly. Insufficient mixing of the ingredients of concrete will result in a reduction in the compressive strength [10].

workability The addition of water to increase workability results in a reduction of the compressive strength of the concrete [11].

---

[10]https://www.ir.nctu.edu.tw/bitstream/11536/29940/1/000167768400013.pdf
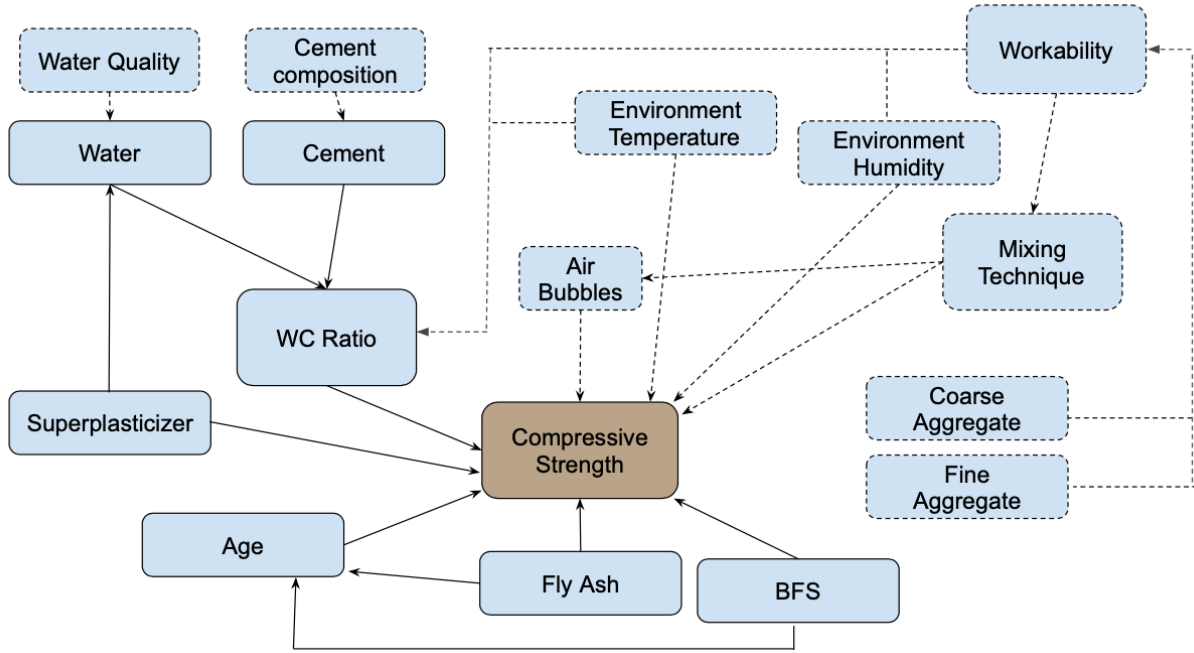[11]https://www.globalgilson.com/blog/what-is-workability-of-concrete

Figure 7: Factors affecting Compressive Strength of Concrete

Air entrainment or air bubbles Small air bubbles are dispersed in the concrete. The effects of air on the compressive strength depend on the bubbles' size and their distribution. The presence of large bubbles reduces compressive strength. Microscopic air bubbles positively affect concrete strength, particularly in cold weather.

6   Conclusion

This study aimed to understand the relationship between different parameters involved in preparing concrete and determining its compressive strength. This understanding can assist in designing concrete mixtures for the desired application reducing costs. We sourced a database with results from testing concrete samples with a wide range of characteristics. We considered the water to cement ratio and the age of the concrete as the independent variables and the compressive strength of concrete as the dependent variable. We adjusted several models to assess the explanatory power of these variables. The impact of adding other ingredients in the binder to improve the strength of the concrete was introduced in the model as categorical variables. We observed that the compressive strength of the concrete has a strong relationship with the water to cement ratio and age. Our results are consistent with the physics of the concrete curing process and can assist in designing concrete recipes to meet desired requirements.