

## CONCRETE STRENGTH

### RELEVANT JMP PLATFORMS AND STATISTICAL TECHNIQUES

Distribution : Histograms, Boxplots, Summary Statistics, Tolerance Interval

Graph Builder : Scatterplots

Multivariate : Correlation, Scatterplot Matrix Plot

Fit Y by X : One variable Linear Regression

Fit Model : Multiple Linear Regression, Profiler, Monte Carlo Simulation

### PROBLEM STATEMENT

Concrete is an essential material used in almost all types of construction. The ability to form concrete into myriad shapes and sizes with different levels of durability makes it an ideal material to use for many different applications such as a simple foundation for a small house all the way to the construction of bridges and large city buildings. Each application requires the concrete used to meet critical mechanical requirements with some cases needing it to support massive amount of weight and sustain heavy vibration such as that from an earthquake.

Manufacturers of concrete perform frequent testing to ensure that each formulation they are creating for a specific application meets critical strength requirements. One such test – the Concrete Compressive strength test – applies a gradual increasing load onto a concrete cylinder to determine the Maximum Compressive Stress the cylinder can sustain before fracture (Figure 1A-C).

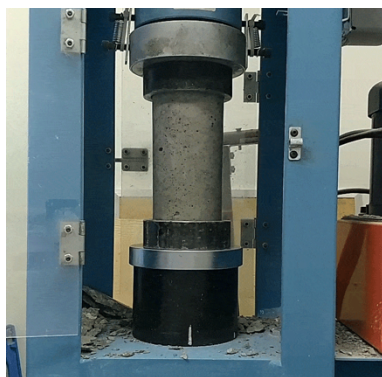


Figure 1A

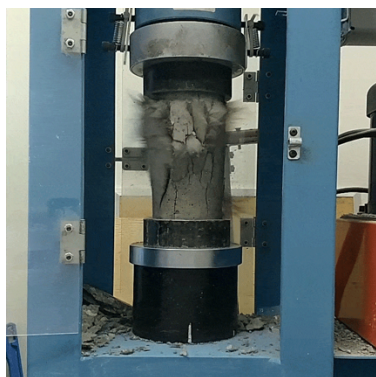


Figure 1B



Figure 1C

Concrete is comprised of four primary ingredients (cement, fine and coarse aggregates, and water) along with the addition of other materials such as fly ash and blast furnace slag which are designed to improve its mechanical and chemical properties.

As a means to improve upon the understanding of how these different ingredients interact influencing the concrete's strength, a dataset was compiled. This dataset consists of test results from a sample of 122 unique batches of concrete that had been made from many different manufacturers across the U.S. The amount of Cement, Coarse and Fine Aggregate, Water, Fly Ash, and Blast Furnace Slag used in each batch was provided. For each batch, the manufacturer had prepared and tested between 2-5 standardized concrete cylinders. The test results provided by these manufacturers is the Average Compressive Stress across the set of cylinders that were tested from each batch.

## DATA SET

Concrete\_Strength.jmp

Strength	The average of the maximum compressive stress (in Mega Pascals MPa) across the set of cylinders tested from each batch.
Cement	Amount (in kg/m <sup>3</sup> ) used in the batch.
Coarse Aggregate	Amount (in kg/m <sup>3</sup> ) used in the batch.
Fine Aggregate	Amount (in kg/m <sup>3</sup> ) used in the batch.
Water	Amount (in kg/m <sup>3</sup> ) used in the batch.
Fly Ash	Amount (in kg/m <sup>3</sup> ) used in the batch. <i>Fly ash is a by product produced in coal-fire power plants that develops cementitious properties when mixed with cement and water.</i>
Blast Furnance Slag	Amount (in kg/m <sup>3</sup> ) used in the batch. <i>Blast Furnance Slag is a by product produced in the making of steel that develops cementitious properties when mixed with cement and water.</i>

## EXERCISES

1. As a means to get familiar with the data, create histograms, boxplots, and summary statistics for each of the variables. *Hint: Use Distribution platform.*
2. Create a Scatterplot Matrix using the 7 variables in order to examine the data via all the bivariate relationships (note: there are 21 pairs).
3. Create a set scatterplots of Strength vs. each of the 6 ingredient variables.  
*Tip: create a scatterplot of Strength vs. Cement using the Graph Builder. Select Redo > Column Switcher under the red triangle setting it up so you can select any ingredient to use as the X variable.*

- a. Choose to add the 'Line of Fit' , 'Confidence Interval' , 'R-Squared' , and 'Equation' to the graph
  - b. Examine each scatterplot to identify how the strength of the concrete changes based upon the amount of each ingredient. Provide a few brief sentences describing those relationships.
4. Perform a single variable linear regression for Strength by each of the ingredient variables  
*Tip: Use Fit Y by X platform.*
  - a. Examine the results of the tests  $H_0 : \beta_1 = 0$  ,  $H_A : \beta_1 \neq 0$  for the model  $Y = \beta_1 + \beta_1 X + \varepsilon$  ,  $\varepsilon \sim N(0, \sigma^2)$ . Identify the ingredient variables that have statistical evidence to indicate a non-zero value for  $\beta_1$  and those do not. Provide a few brief sentences describing your conclusions. If any are non-significant, explain why may it not be correct to conclude that variable has no impact at all on the strength of concrete?
  - b. Is the model  $Y = \beta_1 + \beta_1 X + \varepsilon$  ,  $\varepsilon \sim N(0, \sigma^2)$  being considered for each ingredient reasonable, or should a more complex single variable models be considered? Why is it not a good idea to extrapolate these models much beyond the levels of the ingredients in these data?
5. Build a multivariable regression model that will include terms considered statistically significant. *Tip: Use Fit Model platform.*
  - a. Begin by building a model that contains all possible main effects and 2-variable interactions (i.e., full model). *Hint: Select all the ingredient variables and choose Macro > Factorial to Degree, with the degree value set at 2.*
  - b. Examine the statistical tests shown in the Effects Summary table. Reduce the model using the following strategy. 1) Remove (one term at a time), all the 2-variable interactions not statistically significant until left with only significant ones. 2) Then examine the main effects, remove any non-significant terms provided that they are not involved in any significant 2-variable interaction (i.e., final model).
  - c. Display the 'Actual by Predicted' plot. Comment on how well this final model does at predicting strength.
  - d. Display the 'Residual by Predicted' plot. Are there any extreme values to be concerned with? Is the assumption of equal variance across the range of data appropriate?
  - e. Display the 'Residual by Normal Quantiles' plot. Is the assumption of normal distribution model for the error term in the model appropriate?
6. Perform the following to interpret and implement the final model.
  - a. Provide a brief interpretation of the RSquare and the Root Mean Square Error values.
  - b. Display the equation for this model. *Hint: Select Estimates > Show Predicted Expression under the top red triangle.* Explain why it's difficult to describe, using a single value, how the estimated Strength changes as the amount of each ingredient changes?

- c. Display the Profiler. Using this profiler, find the amount of each ingredient that maximizes Strength. Why may it not be correct to conclude that manufacturers should use this blend of the ingredients to produce concrete with the highest strength?
  - d. Use the prediction profiler to estimate the resulting strength of the concrete with the following ingredient blend: Cement = 300 ; Coarse Aggregate = 950 ; Fine Aggregate 700 ; Water = 200 ; Blast Furnance Slag = 100 ; Fly Ash = 125. Provide a brief sentence description of this estimate including interpretation of the resulting confidence interval. If a manufacturer used this blend to produce concrete, why is it not correct to conclude that all the concrete produced would achieve this estimated strength? Hint: approximately how much of the concrete will have a strength above this estimated value?
  - e. Turn on the Extrapolation Control for this profiler. Describe what this controller is doing. Was it reasonable to use the profiler to estimate strength with the ingredient blend above?
  - f. For this ingredient blend, simulate the Strength of 10000 batches of concrete. Hint: Select 'Simulate' under the Profiler Red Triangle. Choose Fixed values for all the ingredients, but 'Add Random Noise' for the Response.
  - g. Estimate the strength that 99% of concrete produced at this ingredient blend will achieve. Hint: Run the Distribution script created in the new simulated data table. Select 'Tolerance Interval' under the Profiler Red Triangle. Enter 0.99 as the 'Proportion to cover'. Select 'One-sided lower limit' and
7. What other information would it have been useful to have in this dataset that was unfortunately not included? What else would be needed in order for us to create a statistical model that describes almost all the variation in concrete strength produced at a given ingredient blend (i.e., a model with a very high RSquare)?