JMP Instructional data set 032

Concrete Strength

JMP STATISTICAL DISCOVERY

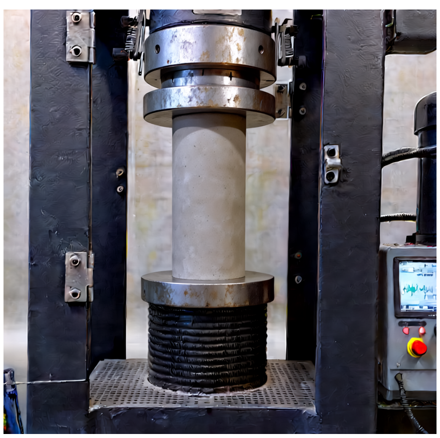
JMP platforms and statistical methods

Distribution:	Histogram; box plots; summary statistics
Graph Builder:	Scatter plots; simple linear regression
Multivariate:	Correlation; scatterplot matrix
Fit Model:	Multiple variable regression

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 that the concrete used meets critical mechanical requirements, and in some cases, needing it to support massive amount of weight and sustain heavy vibration, such as that from an earthquake.

Concrete manufacturers 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.



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.

To improve upon the understanding of how these different ingredients interact and influence the concrete's strength, data was collected. This data set 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. The data values provided by these manufacturers are the average maximum compressive stress across a set of cylinders tested.

Data

IDS_032.jmp

Strength	Average maximum compressive stress (in Mega Pascals, MPa) across a set of cylinders tested
Cement	Amount (kg/m^3) used in each batch
Coarse Aggregate	Amount (kg/m^3) used in each batch
Fine Aggregate	Amount (kg/m^3) used in each batch
Water	Amount (kg/m^3) used in each batch
Fly Ash	Amount (kg/m^3) used in each batch *Fly ash is a byproduct produced in coal-fired power plants that develops cementitious properties when mixed with cement and water*
Blast Furnace Slag	Amount (kg/m^3) used in each batch *Blast Furnace Slag is a byproduct produced in the making of steel that develops cementitious properties when mixed with cement and water.*

Exercises

1. Create histograms, boxplots, and summary statistics for all the variables to get familiar with the data.

   Instructions:
   Analyze > Distribution. Place all variables in the Y, Columns. Click OK.

2. Create a Scatterplot Matrix using all seven variables to examine the data via. all the bivariate relationships. Note: there are 21 pairs.

   Instructions:
   Analyze > Multivariate Methods > Multivariate. Place all variables in the Y, Columns. Click OK.

3. Create scatterplots of Strength vs. each of the six ingredient variables. Examine each variable's relationship with Strength. Comment on those that appear to have a relationship with Strength and those that do not. Are there any data points that stand out as being much different than the others?

   Instructions:
   Graph > Graph Builder. Place 'Strength' on the Y axis. Place 'Cement' on the X axis. Select Redo > Column Switcher under the red triangle. Select 'Cement' as the Column to Switch, and then select that along with all the other variables for the Replacement Columns. Click OK. Now the variable to plot on the X axis can be chosen in the Column Switcher controls on the left.

4. Add linear regression lines to the graph created in Exercise 3, along with the R-Squared values and an F Test testing for the significant of a non-zero slope (at 0.05 significance level). Comment on those that are statistically significant from those that are not. Are these results consistent with what you concluded in Exercise 3?

> ***Instructions:***
> *Choose the Line of Fit icon in the graph palette.* 
>
> *Select the check boxes for R-Squared, Equation, and F Test in the controls on the left.*

5. Create a multivariable linear regression model that includes all main effects and two-factor interaction terms.

> ***Instructions:***
> *Analyze > Fit Model. Place 'Strength' in the Y role. Select all the other variables in the Columns list on the left. Then select Macros > Factorial to Degree, ensuring that "2" is the value in the Degree field. Choose Minimal Report in the Emphasis drop down menu. This can be referred to as the "full model".*

6. 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 two-factor interactions not statistically significant (at the 0.05 level) until left with only the significant ones. 2) Examine the statistical tests for the main effects. Remove (one term at a time), all the non-statistically significant main effects, but not removing any that are included in any significant two-factor interaction.

> ***Instructions:***
> *Highlight a term in the Effects Summary table, and select Remove.*
> *The final model should only include two-factors interactions with p-values < 0.05, and main effects with p-values < 0.05 and ones that are part of a significant two-factor interaction. This is often referred to as the final reduced model.*

7. Evaluate the assumptions of equal variance and normality of the residual in the final model. Is there any cause for concern? Return to the scatter plots created in Exercise 3. Are any of the data values you considered unusual in those graphs unusual with respect this final model? Why might that be?

> ***Instructions:***
> *Choose Row Diagnostics > Plot Residual by Predicted and Row Diagnostics > Plot Residual by Normal Quantiles under the red triangle.*
>
> *In the scatterplots in Exercise 3, highlight any data value that you considered to stand out as being unusual. Return to the Residual by Predicted Plot and Residual Normal Quantile plot. That data value will also be highlighted in those plots.*

8. How well do these variables, through the model, explain/predict the strength of the concrete? Include quantitative values in your description. What are some potential causes of the remaining variation that the model is not able to account for?

> ***Instructions:***
> *Choose Row Diagnostics > Plot Actual by Predicted under the red triangle.*
> *Examine R-Squared and Root Mean Square Error (RSME).*

9. Display the equation for the model. Why is it not possible to use the coefficients for the main effects to describe how strength changes as the amount of that individual ingredient changes?

> ***Instructions:***
> *Choose Estimates > Show Predicted Expression under the red triangle.*

10. Display the Prediction Profiler for the model. Use this tool to predict the Strength of a batch of concrete made under the two following blend of ingredients. Report both the estimated expected value and 95% confidence interval. Do you trust the estimates for both of these blends? Explain why or why not.

    a) Cement = 226 , Coarse Aggregate = 975 , Fine Aggregate = 739 , Water = 183 , Blasts Furnace Slag = 115 , Fly Ash = 119

    b) Cement = 400 , Coarse Aggregate = 1000 , Fine Aggregate = 850 , Water = 220 , Blasts Furnace Slag = 175 , Fly Ash = 70

> ***Instructions:***
> *Choose Factor Profiling > Profiler under the red triangle.*
> *Type in the values for each ingredient for each blend. The estimated expected value and confidence interval is displayed on the left*
>
> *Uder the red triangle for the Prediction Profiler, turn on the Extrapolation controls.*
> *Try to enter the same values as above for each blend of ingredients.*
> *To see how the extrapolation controls are being applied to the profiler, manually drag the red dotted line for the amount of an ingredient and watch how the acceptable range for the amounts of the other ingredients change.*

11. Find the blend of the ingredients that maximize Strength.

> ***Instructions:***
> *Choose Optimization and Desirability > Maximize Desirability under the red triangle for the Prediction Profiler.*

12. What other data could be collected and/or experiments run to develop a more thorough understanding of how the various ingredients impact the strength of the concrete?