

Practice: Fitting Polynomial Models

In JMP, open the file **VSSTeamData.jmp**. Make sure that the five outliers for **Yield** are hidden and excluded.

1. Using Graph Builder, create a scatterplot for **CI** (in the Y zone) and **Xf** (in the X zone). Use the slider for the smoother to explore the relationship.

How would you describe the relationship between the two variables?

The overall relationship is curvilinear.

2. Using **Fit Y by X**, fit a linear regression model for **CI** (as **Y, Response**) and **Xf** (as **X, Factor**). To fit the regression model, select **Fit Line** from the red triangle next to **Bivariate Fit**. Select **Plot Residuals** from the red triangle next to **Linear Fit**.

Does a linear model describe the relationship between **CI** and **Xf**?

No, there is curvature in the relationship, and you can see this curvature in the residuals. This indicates that a linear model does not make sense.

3. Use the **Fit Polynomial** menu under the top red triangle to fit a quadratic model for **CI** and **Xf**. Select **Plot Residuals** from the red triangle next to **Polynomial Fit Degree=2**.
4. Look at the Parameter Estimates and the Summary of Fit tables for the linear and quadratic models, and look at the residual plots.

Which model does a better job of describing the relationship between **CI** and **Xf**?

The quadratic model does a better job of explaining the relationship between the two variables. The residuals for the quadratic model don't show curvature, and the squared term in the quadratic model is significant. RMSE is much lower for the quadratic model, and RSquare and RSquare Adjusted are higher.

5. Use the **Fit Polynomial** menu under the top red triangle to fit a cubic model for **CI** and **Xf**. Select **Plot Residuals** from the red triangle next to **Polynomial Fit Degree=3**. Compare the linear, quadratic, and cubic models.

Which model does the best job of explaining the variation in **CI**? Which model is best at describing the relationship between **CI** and **Xf**?

The RSquare and RSquare Adjusted statistics for the cubic model are higher, the cubic model has the lowest RMSE, and the cubic term in the parameter estimates table is significant. Therefore, the cubic model is best at describing the relationship.

Hide Solution

Statistical Thinking for Industrial Problem Solving

Copyright © 2020 SAS Institute Inc., Cary, NC, USA. All rights reserved.

Close