

# Meeting 8

PLS-PM Cross-validation and Simulated Data

04/27/2015

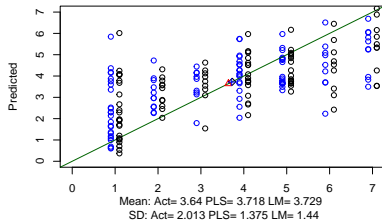
# Cross-validation Method

This is the process performed to produce the Cross-validation Samples:

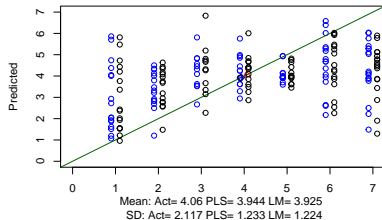
- ▶ Load the original 183 observations.
- ▶ Select a random sample of 183 observations from the original data (Randomize the Observations).
- ▶ The first 100 randomized observations were divided into 10 subsamples of 10 observations each one.
- ▶ Each subset of 10 observations was taken as a holdout for an individual model, and the remaining 173 observations as a training data.
- ▶ Each model produced its own predictions.
- ▶ All the predictions were combined together to produce a set of 100 predictions based on 10 different models, with different training and holdout data.

# Actual vs Predicted Scatterplot (PLS vs LM)

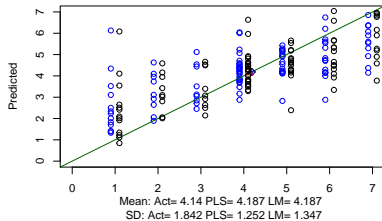
Actual vs Predicted (AA.0)



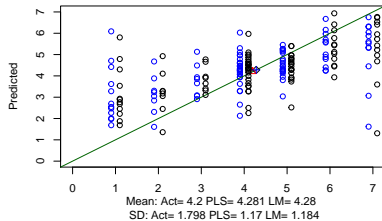
Actual vs Predicted (AA.1)



Actual vs Predicted (AA.2)

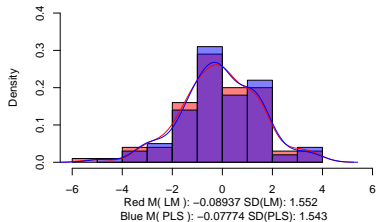


Actual vs Predicted (AA.3)

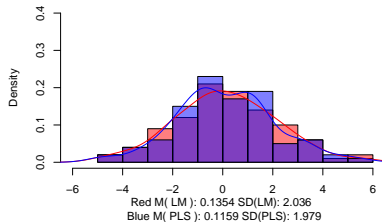


# PLS vs Linear Regression: Joint Histogram

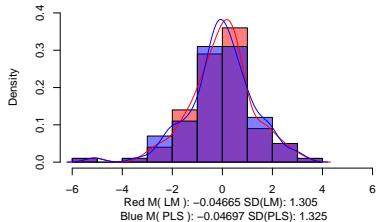
**PLS vs LM Residuals AA.0**



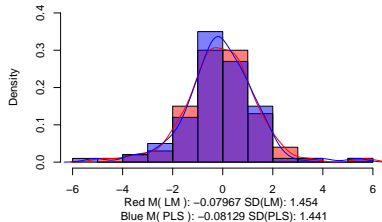
**PLS vs LM Residuals AA.1**



**PLS vs LM Residuals AA.2**

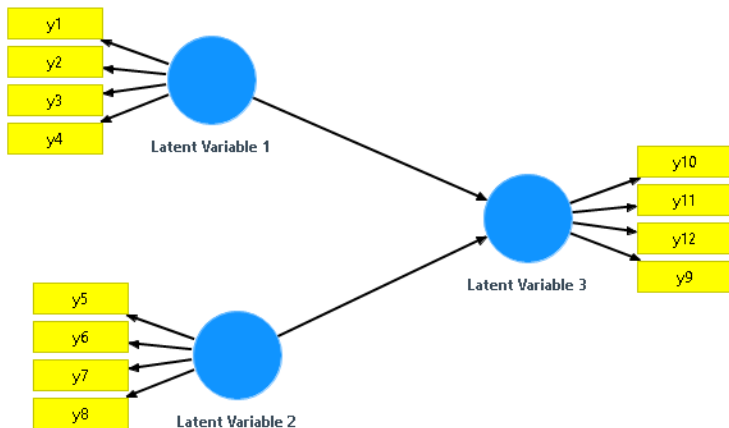


**PLS vs LM Residuals AA.3**



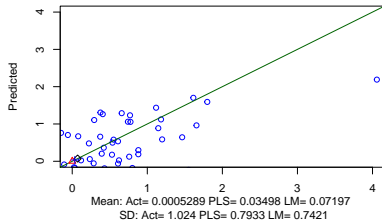
# Simulated Data Model

Suneel prepared simulated data which resulted in the following model:

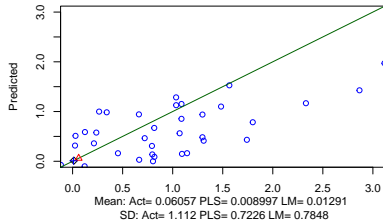


# Actual vs Predicted Scatterplot (Simulated Data)

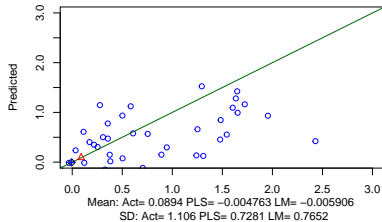
**Actual vs Predicted (y9)**



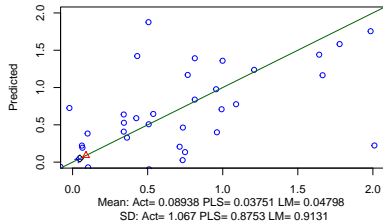
**Actual vs Predicted (y10)**



**Actual vs Predicted (y11)**

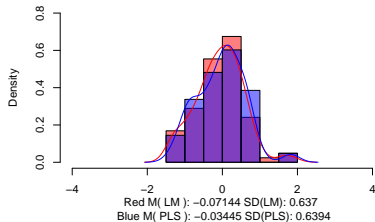


**Actual vs Predicted (y12)**

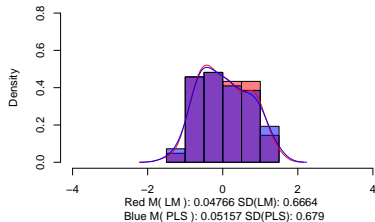


# PLS Residuals (Simulated Data)

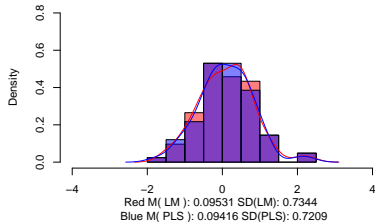
**PLS vs LM Residuals y9**



**PLS vs LM Residuals y10**



**PLS vs LM Residuals y11**



**PLS vs LM Residuals y12**

