MORS

Form 712A Non-Government Disclosure (Contractor, Academic, UARC etc.)

Check the Event Website for the form deadline at www.mors.org.
If using the MORS Presenter Center all forms must be upload in Presenter Center. For other events e-mail the completed form to 712s@mors.org. If the presentation classification and or Disclosure Statement below is different than those entered in the MORS Presenter Center, correct the Presenter Center entries to ensure proper labeling of the Presentation.

PART I

Author Request - The following author(s) request authority to disclose the following presentation at the MORS Event below with subsequent publication in the MORS Event Report and posting on the MORS website if applicable.

Principal Author **Thomas A. Donnelly**	Other Author(s)
Principal Author's Organization and complete mailing address **SAS Institute Inc.** **27 Farmingdale Ln** **Newark, DE 19711**	Principle Author Phone **302-489-9291** · Principle Author FAX **919-677-4444** Principle Author E-mail **tom.donnelly@jmp.com**
Principal Author's Signature X *Thomas A. Donnelly*	Date **1 June 2018**
MORS Event **86th MORS Symposium**	Event Date(s) **18-21 June 2018**

Presentation Type ☐ Plenary ☐ Course ☑ Tutorial ☐ Special Session ☐ Poster ☐ Demonstration

☐ Working/Composite/Distributed or Focus Group List All _____ ☐ Other _____

Title of Presentation Exploratory Data Analysis and Root Cause Analysis - "What Can You Do When You Don't Have a Designed Experiment?"	Presentation ID (if assigned) **38552**

Classification ☐ SECRET ☐ SECRET//REL TO FVEY ☐ CONFIDENTIAL ☐ CONFIDENTIAL//REL TO FVEY

☑ UNCLASSIFIED ☐ UNCLASSIFIED W/FOUO ☐ Other _____

Distribution Statement ☑ A (Publicly Releasable) ☐ B ☐ C ☐ D ☐ E (see side 2 for definitions)

A.	This work was performed in connection with a government contract.	☐ YES (Complete Parts I, II & III)
B.	This presentation is based on material developed by the author as part of company/organization approved research e.g. IR&D and was NOT done under a government contract.	☐ YES (Complete Parts I, II & III)
C.	This presentation was NOT done under a government contract, contains no government information, is my own work and is approved for public release.	☑ YES (Complete Part I only)

EXPLORATORY DATA ANALYSIS AND ROOT CAUSE ANALYSIS

"What can you do when you don't have a designed experiment?"

86th MORSS at NPS
Monterey, CA
June 18, 2018

Tom Donnelly, PhD, CAP
Sr. Systems Engineer
& Co-insurrectionist
JMP Federal Government Team
tom.donnelly@jmp.com

TOPICS

Exploratory Data Analysis and Root Cause Analysis

- Plot the data

  - Dynamic linking – look for correlation between behavior and possible causes – make lots of graphs

    - Just don't think that **correlation** necessarily means **causation**!

  - Use data filters to isolate factors and levels

  - Preparing the data – outliers, missing, recoding

  - Example data sets

    - Pharmaceutical – continuous & binary – root cause

    - Breast cancer – binary - PCA

    - Cyber Attack – Multinomial - binary

    - Diamonds – lots of categorical classifications

    - FAA Data – daily records

    - Positional – latitude & longitude (GPS, satellite, image)

      - GPS, SST, Drone

    - What Type of Data Do You Have?

jmp

§sas | THE POWER TO KNOW.

Exploratory Data Analysis and Root Cause Analysis

- Just how much are my factors correlated?

  - Matrix – VSS cont/cat response

  - PCA – breast cancer

- Create a design from candidate set –

  - Phamaceutical Tablet Data

  - Diamonds Data

  - FAA Data

- Use Data Mining Tools

  - Use a Robust "Honest Assessment Method" – Train, Validate & Test

  - Other Cross-Validation approaches – K-fold, Leave-One-Out

  - Penalization criteria – BIC, AICc, ERIC, Adjusted R-Square

Exploratory Data Analysis and Root Cause Analysis

- Decision tree methods (Useful when too many columns for regression)

- Regression – Linear, Generalized Linear (logistic), Stepwise (not always), Penalized (GenReg)

- Neural

- CKPS Today? Tomorrow? Physics (timeless) vs People (shifty/adaptive)

- Case Study - FAA Data/Models

  - Cross-Validation approaches – K-fold, Leave-One-Out

  - Penalization criteria – BIC, AICc, ERIC, Adjusted R-Square

INTERACTIVE SELECTION

LOOK FOR CORRELATIONS BETWEEN MEASURED RESPONSES AND POTENTIAL FACTORS
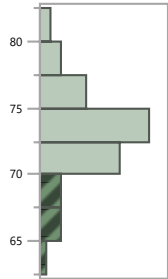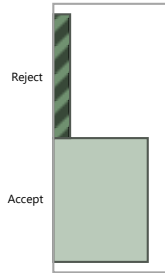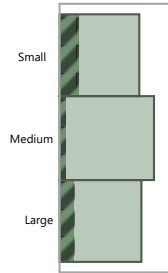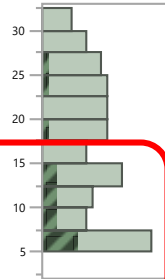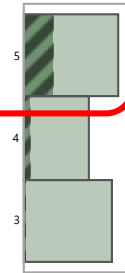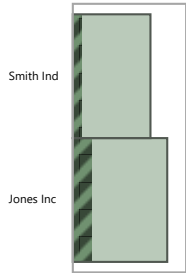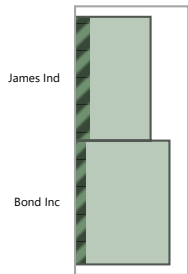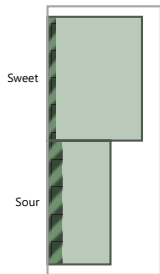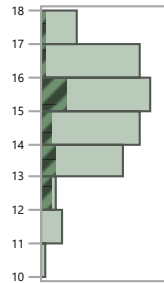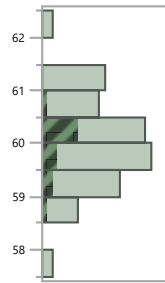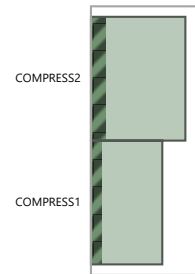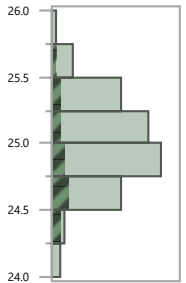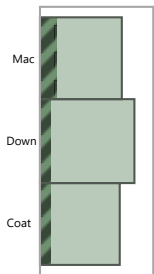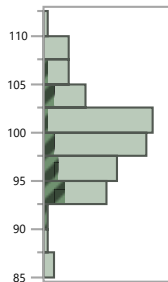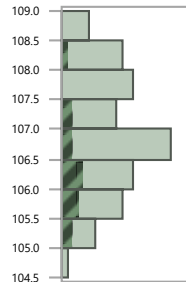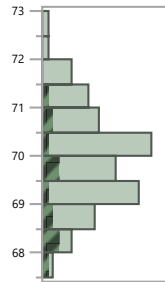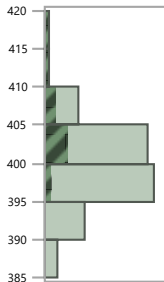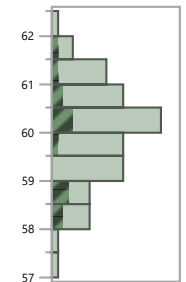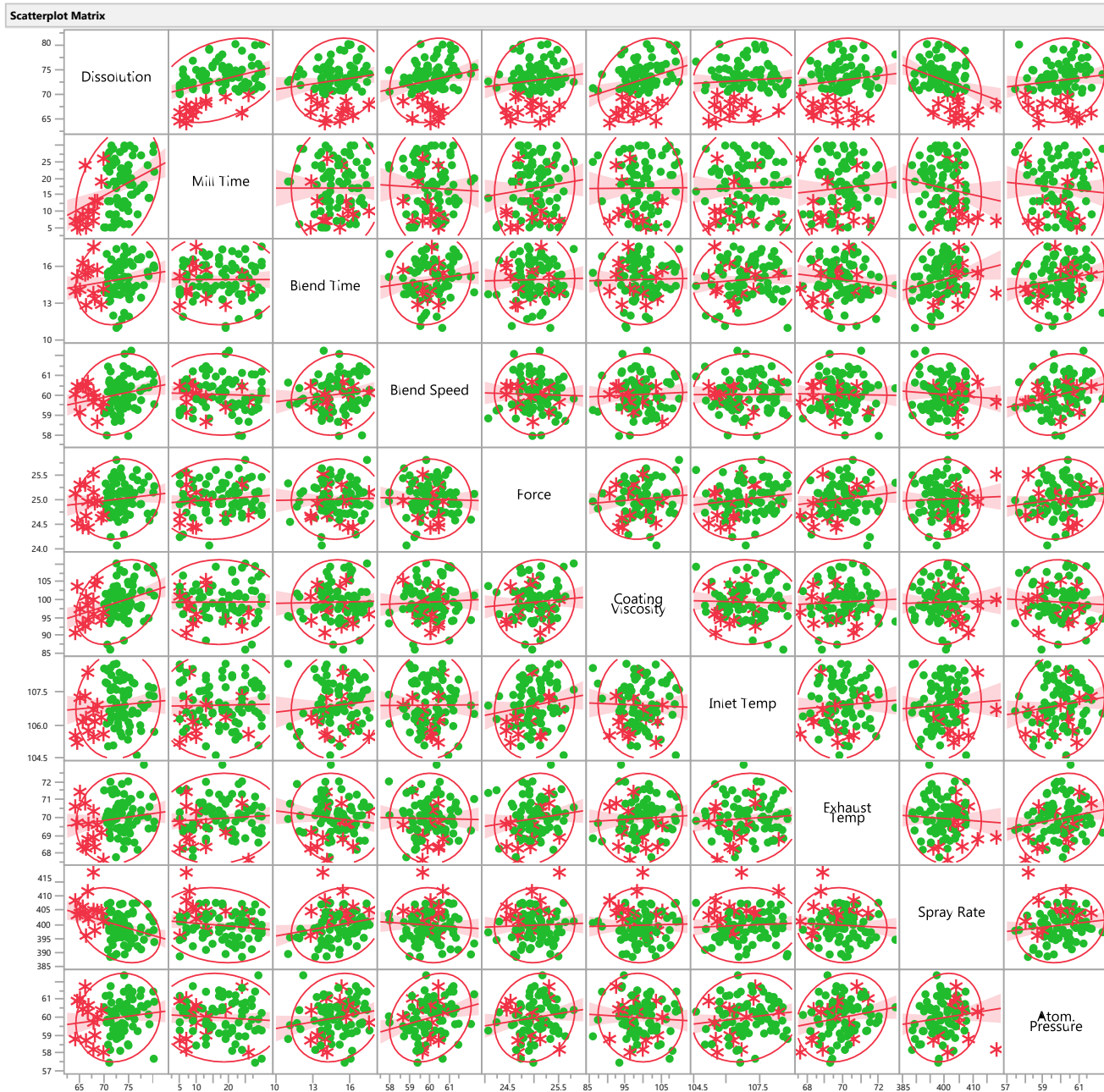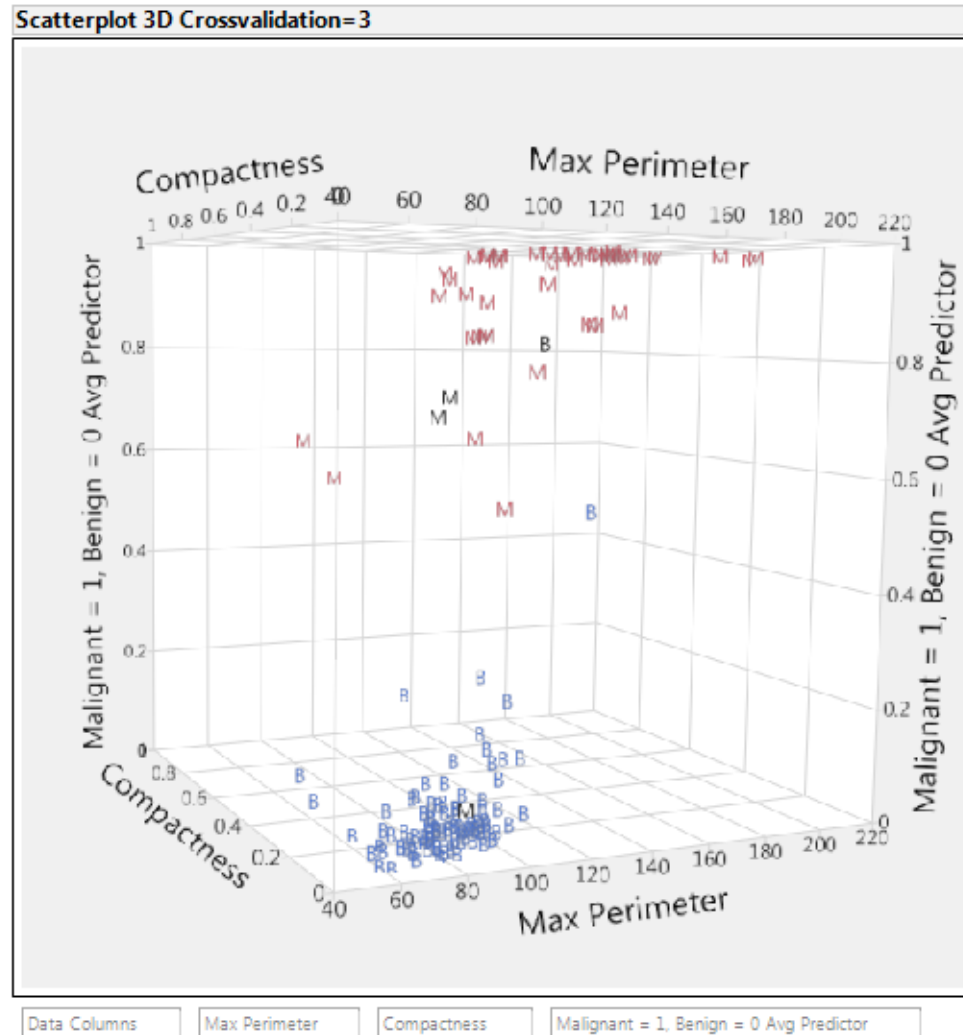
CORRELATION OF CONTINUOUS COLUMNS

Look for correlations between responses and potential factors, as well as factors with each other

Scatterplot Matrix

ACTUAL DIAGNOSIS (M OR B) PROBABILITY OF MALIGNANT AND TWO MAIN FACTORS

Thanks.
Questions or comments?

TOM.DONNELLY@JMP.COM