Predictive Analytics: *Employing the Disruptive Behavior Reporting System (DBRS)*
Battle Creek VA Medical Center

Jamie Cole, Health Informatics Intern
Health Informatics and Bioinformatics Graduate Program
Grand Valley State University, Allendale, MI, USA

Building a Predictive Model

Select Necessary (and Available) Variables for Answering the Research Questions:

- Determine level of measurement (nominal, ordinal, interval)
  - If interval, determine if the variable is continuous or discrete
- Level of measurement determines the type of information producible

The **DBRS** is a secure, web-based, and facility-managed data collection site of disruptive behavior incidents transpiring at VHA facilities.

The system allows authorized individuals to receive, review, maintain and document disruptive behaviors.

Research Questions: What variables contribute to disruptive behavior incidents occurring at VHA facilities? How do these significant variables interact with each other?
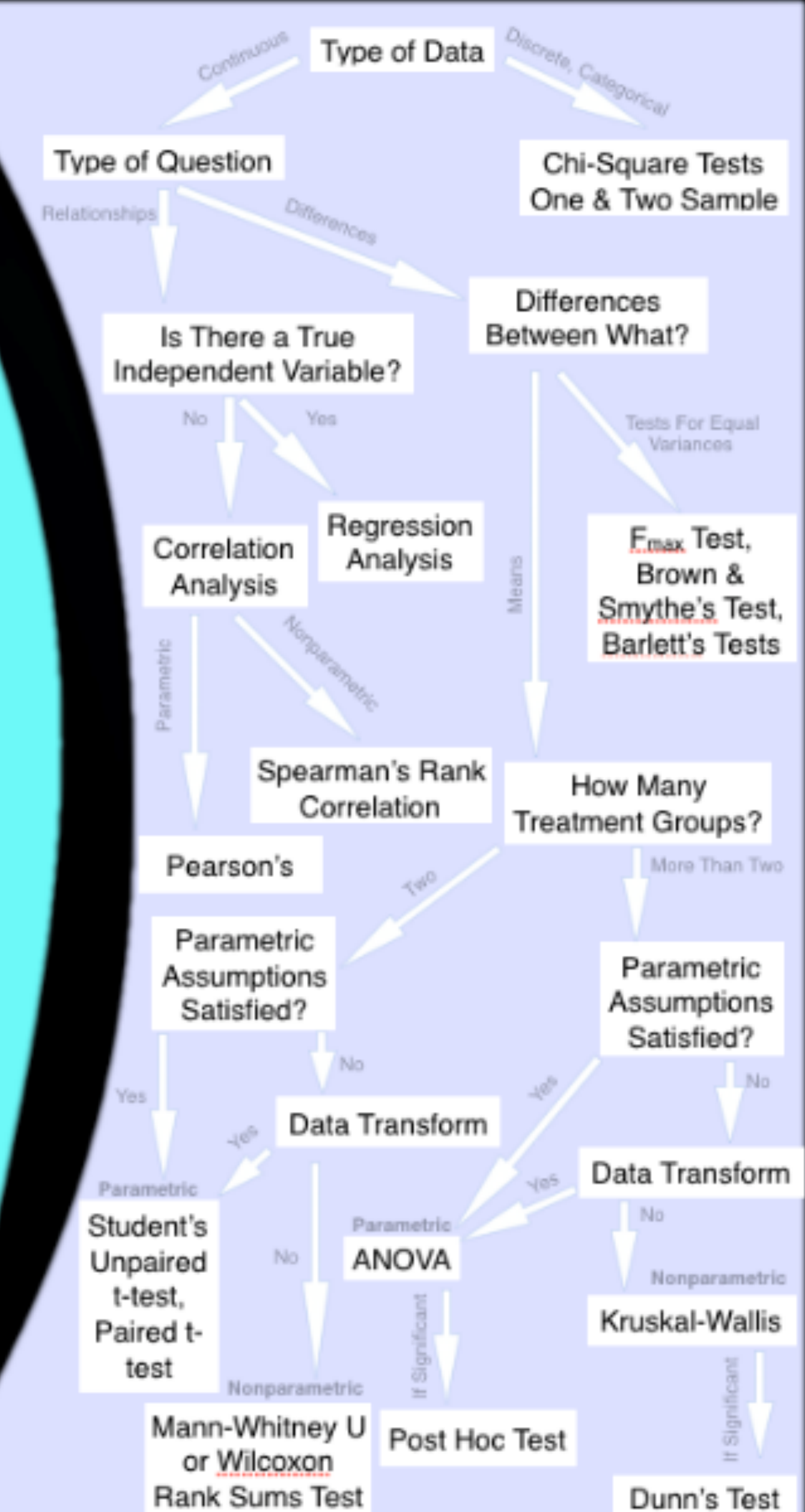
Variables

Incident
- Status
- Shift
- Tour · Role
- Engager · Station
- Behaviors · Interventions
- Event Date · Incident Class
- Conditions · Reference Number
- Restraints
- Other information · Experiencer
- Location Code · Occupation
- Severity of Outcome · WBRA Location
- Suspected Sexual Assault · Related to Drugs or Alcohol
- Confirmed Sexual Assault
- Criminal or purposely Unsafe Act
- Disruptive Behavior (Physical)
- Disruptive Behavior (Verbal)
- Patient Abuse

Individual
- Flag · Meaningful Social Supports
- Sleeping Problems · Military Sexual Trauma
- Chronic Pain · DOB
- Combat Vet Status · Unstable Relationships with Others
- Alcohol Misuse · Self-Care
- Probably PTSD + Anger · Diabetes
- Gender · Major Destabilizing Behaviors,
- Financial Stability · Moods, Thoughts
- Stable Housing · Resilience
- Compliance with Treatment · Branch
- Substance Abuse Problems
- Work/Volunteer/Vocation/Education
- Perceived Self-Determination
- Combat Vet Status
- Sleeping Problems
- Basic Needs Met
- Chronic Pain
- Isolation

Check for and Resolve Data Issues That Can Affect the Model:

- Missing data
- Multicollinearity
- Truncation and censoring
- Outliers and influential points

Collect, Code, Enter, and Clean Data

(Statistical decision flowchart: Type of Data — Continuous / Discrete, Categorical; Type of Question; Relationships / Differences; Chi-Square Tests One & Two Sample; Is There a True Independent Variable?; Differences Between What?; Correlation Analysis; Regression Analysis; Tests For Equal Variances; F_max Test, Brown & Smythe's Test, Bartlett's Tests; Spearman's Rank Correlation; How Many Treatment Groups?; Pearson's; Parametric Assumptions Satisfied?; Parametric Assumptions Satisfied?; Data Transform; Student's Unpaired t-test, Paired t-test; ANOVA; Data Transform; Mann-Whitney U or Wilcoxon Rank Sums Test; Post Hoc Test; Kruskal-Wallis; Dunn's Test)

Run Several Versions of the Model

- Univariate & bivariate statistics
- Check for over-dispersion
- Test (possibly eliminate) interactions
- Place predictors in sets
- Explore other types of non-linearity
- Drop nonsignificant control variables
- Test for the best specification of random effects
- Do hierarchical modeling to see the effects of predictors added in blocks or alone

Compare Models

- Control variables vs those without
- Statistical Significance of Parameters
- Coefficients
- R-squared
- Error term
- Adjusted r-square
- Mallow's Cp criterion
- Akaike information criterion (AIC)
- Bayesian information criterion (BIC)

Select the Best Model:

Cross-Validation,
Regression Regularization,
Goodness of fit

- Hard to find any relationships when everything is put in all at once
- Place predictors into theoretically distinct sets
  - Variables within a set are often correlated
  - See how related variables work together
- Build each set separately → theoretically meaningful models → solid understanding of how the pieces fit together